DOCUMENT RESUME

ED 027 281 24 SP 002 422

By-Brown, Bob Burton

An Investigation of Observer-Judge Ratings of Teacher Competence. Final Report.

Florida Univ., Gainesville.

Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.

Bureau No-BR-5-1073

Pub Date 31 Jan 69

Contract-OEC-6-10-288

Note-148p.

EDRS Price MF-$0.74 HC-$7.50

Descriptors-Analysis of Variance, Attitudes, Attitude Tests, Behavior Rating Scales, Check Lists, Correlation, Educational Experiments, *Evaluation, Factor Analysis, Interaction Process Analysis, Interinstitutional Cooperation, *Lesson Observation Criteria, *Student Teachers, *Teacher Certification, *Teacher Qualifications

Identifiers-CBRS, Classroom Behavior Rating Scale, Dogmatism Scale, D-Scale, PBI, Personal Beliefs Inventory, Teacher Evaluation Scale, Teacher Practices Inventory, Teacher Practices Observation Record, TES, TPI, TPOR

Demonstrating and testing Conant's recommendation that teacher competence should determine certification, this four-phase study from 1964-68 developed and field-tested procedures for evaluating teacher competence and for determining how evaluation is affected by the beliefs of student teachers and observer-judges. Following the Phase I orientation of observer-judges and evaluation of rating instruments and procedures, 539 observer-judges from colleges, public schools, and State Departments of Public Instruction rated 407 student teachers' clinical classroom performances over a one-year period (Phase II) with Teacher's Classroom Behavior instruments. Prior to rating, students and observer-judges took three Study of Beliefs tests. Phase II data was statistically analyzed and compared with data from Phase III in which 100 Phase II subjects, then first-year teachers, and 100 experienced teachers were evaluated. Data analysis in Phase IV revealed: predictable interrelationships among teacher beliefs, teacher competence, observer descriptions, and observer-judge beliefs; belief gaps between colleges of education and public schools; and theory-practice discrepancies in teachers and observer-judges. Recommendations for use of these findings in teacher evaluation programs are made. (LP)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED0 27281

BR 5-1073
PA-24
BR OE

Final Report

Project No. D-182
Contract No. OE-6-10-288

AN INVESTIGATION OF OBSERVER-JUDGE RATINGS OF TEACHER COMPETENCE

Bob Burton Brown
University of Florida
Gainesville, Florida
January 31, 1969

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

SP002422

ACKNOWLEDGMENTS

Sincere appreciation is expressed to the many teachers and observer-judges whose cooperation made possible the collection of data for this study. For their assistance in organizing and supervising the collection of the data at the various teacher education institutions, we are indebted to Dr. Leonard Kennedy, Dr. James Cochrane, Dr. Elizabeth Jalbert, Dr. Herbert Kliebard, Dr. George Beauchamp, Dr. Robert Maidment, Dr. Peter Oliva, and Mrs. Maude Watkins.

Invaluable contributions to the project were rendered by research assistants, Dr. Jeaninne N. Webb, Dr. Tom R. Vickery, Mr. Winston Summerhill, Dr. Rodney Johnson, Dr. Herschel Shosteck, Dr. Julie Joyce Conry, and Mr. Wm. Nicholas Stoffel.

Appreciation is expressed to those who rendered services in the processing and analysis of the data, Dr. William Mendenhall, Dr. P. V. Rao, Dr. Fred Barnett, Mr. Robert Beaver, Mr. James T. McClave, Dr. Frank Vickers, and Dr. Charles Bridges.

Special contributions as consultants to the project were made by Dr. Norman Bowers, Dr. John Guy Fowlkes, Dr. George Beauchamp, and Dr. Nicholas Fattu.

For their participation in the initial conception and planning of the study, appreciation is due Dr. Merle Borrowman, Dr. Wilson Thiede, Dr. Lindley J. Stiles, Dr. Julian Stanley, Dr. William Cartwright, Dr. John I. Goodlad, Dr. B. J. Chandler, and Dr. James B. Conant.

Finally, appreciation is expressed to Marion Terry and Vida Hamilton, secretaries, Tandel Brown, clerk-typist, and, above all, Marie Smallwood, for her skills as typist, accountant, secretary, coder and keeper of data, coordinator of staff, and loyal right arm.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS ii

LIST OF TABLES AND FIGURES v

CHAPTER

I. INTRODUCTION 1
    The Problem 1
    Objectives 3
    Research Foundation 3

II. RESEARCH PROCEDURES 6
    General Design 6
    Subjects 6
    Instrumentation 7
    Data Collection 8
    Evaluation of Data 9

III. PHASE I - PILOT STUDY 11
    Purposes 11
    Subjects 11
    Data Collection Procedures 11

  Reliability Estimates - Teacher Practices
    Observation Record 12
    Analysis of Data 12
    Findings - TPOR scores 13
    TPOR Reliability Summary 22

  Reliability Estimates - Classroom Behavior Rating
    Scale 22
    Analysis of Data 22
    Findings - CBRS Scores 23
    CBRS Reliability Summary 25

  Comparison of the Teacher Practices Observation
    Record and The Classroom Behavior Rating Scale 26

  Identification of Predictor Variables. 28
    Teacher Practices Observation Record 28
    Classroom Behavior Rating Scale 36

IV. PHASE II - OBSERVER-JUDGE RATINGS OF STUDENT TEACHERS 48
    Purposes 48
    Procedures 48
    Analysis of Data - Using Total Scores of
      Observations and Evaluations 49
    Findings of Total Score Analysis 56
    Discussion of Total Score Analysis 56
    Summary of Total Score Analysis. 62

CHAPTER Page

Re-examination of the Data Using Factor Scores . . 62
  Analysis of Data Using Factor Scores 63
  Findings of the Factor Scores 64
  Discussion of the Factor Scores 66
  Summary of Factor Score Analysis 69

V. PHASE III (FOLLOW-UP STUDY) 71
  Purposes 71
  Procedures 71
  Data Collection 72
  Analysis of Data 72
  Findings 74
  Discussion 74
  Conclusions 81

VI. COMPARISON ANALYSIS OF PHASE II AND PHASE III DATA . 82

VII. SUMMARY, CONCLUSION, AND RECOMMENDATIONS 87
  Summary of Findings 88
  Conclusions 93
  Recommendations 95

APPENDIX A 98

APPENDIX B 110

APPENDIX C 126

LIST OF TABLES

TABLE Page

1. Mean TPOR Scores Given Five Films by All Observers . . . 13

2. Mean TPOR Scores Given Films on Repeated Observations One Year Apart 14

3. Correlation of TPOR Scores Obtained from Repeated Observations of Films 15

4. "Within-Observer" Reliability Coefficients for TPOR Scores on Repeated Viewings of Films 21

5. TPOR Internal Consistency Reliability Coefficients . . . 21

6. Mean CBRS Scores for Films 23

7. "Within-Observer" Reliability Coefficients for CBRS Scores on Repeated Viewings of Films 24

8. CBRST and CBRSP Internal Consistency Reliability Coefficients 25

9. The Relationship Between TPOR Means and Evaluative Ratings of Five Filmed Teaching Episodes 27

10. CBRS Means and Evaluative Ratings 28

11. Definition of Variables 29

12. Interaction Model - TPOR - Summary of F-tests to Determine Significance of Variables in the Model . . . 32

13. Summary of F-tests to Determine Significance of Variables in the Model - TPOR - Order 1 33

14. Summary of F-tests to Determine Significance of Variables in the Model - TPOR - Order 3 34

15. Final Model for TPOR, Analysis of Variance, Summary . . 35

16. Interaction Model - CBRSP - Summary of F-tests to Determine Significance of Variables in the Model . . . 37

17. Summary of F-tests to Determine Significance of Variables in the Model - CBRSP - Order 1 38

18. Summary of F-tests to Determine Significance of Variables in the Model - CBRSP - Order 3 39

TABLE Page

19. Final Model CBRSP, Analysis of Variance, Summary 41

20. Summary of F-tests to Determine Significance of Variables in the Model - CBRST - Order 1 42

21. Summary of F-tests to Determine Significance of Variables in the Model - CBRST - Order 3 43

22. Final Model CBRST, Analysis of Variance, Summary 44

23. Summary of F-tests to Determine Significance of Variables in the Model - CBRS - Order 1 45

24. Summary of F-tests to Determine Significance of Variables in the Model - CBRS - Order 3 46

25. Final Model CBRS, Analysis of Variance, Summary 47

26. Definition of Variables 51

27. Variables Which Contribute to Variance in Observation of Classroom Behavior 57

28. Variables Which Contribute to Variance in Evaluations of Teaching Competence 58

29. F-values of Theoretical and Empirical Factors 63

30. Teacher Practices Observation Record (TPOR) Regression Theoretical Factors 64

31. Teacher Practices Observation Record (TPOR) Regression Empirical Factors 64

32. Teacher Evaluation Scale (TES) Regression Theoretical Factors 65

33. Teacher Evaluation Scale (TES) Regression Empirical Factors 65

34. Variables Which Contribute to Variance in Observation of Classroom Behavior 75

35. Variables Which Contribute to Variance in Evaluations of Teaching Competence 76, 77

36. Results of Paired Difference t-tests Comparing Phases II and III 83

37. Multiple Regression - Model 1 85

38. Multiple Regression - Model 2 86

FIGURE Page

1. Relationship of TPOR and CBRS Mean Scores 26

CHAPTER I

INTRODUCTION

The purpose of this study was to field test the use of judgments of teacher competence in classroom performances as the potential basis for teacher certification. In order to do this the project demonstrated and evaluated a number of ways in which both academic and education professors, supervisors of student teaching, cooperating public school teachers and administrators, and State Department of Public Instruction personnel may be brought together in teams to observe classroom teaching performances and to judge competence for teaching.

The Problem

As a result of a two-year study of teacher education and certification policies, James B. Conant concluded: "The policy of certification based on the completion of state-specified course requirements is bankrupt."[1] Conant pointed out that completion of specific academic and professional courses or programs approved by a national accrediting agency "cannot be enforced in such a manner that the public can be assured of competent teachers, and they involve the states in acrimonious and continuous political struggles, which may not serve the public interest."[2] Consequently, Dr. Conant suggested the need for alternative programs and policies which rely primarily on the use of judgments of competence of classroom teaching performance as the basis for teacher certification.

The first and most central of the 27 recommendations by Conant concerns certification based on evidence of competence:

> For certification purposes the state should require only (a) that a candidate hold a baccalaureate degree from a legitimate college or university, (b) that he submit evidence of having successfully performed as a student under the direction of college and public school personnel in whom the State Department has confidence, and in a practice-teaching situation of which the State Department approves, and (c) that he hold a specially endorsed teaching certificate from a college or university which, in issuing the official document, attests that the institution as a whole considers the person adequately prepared to teach in a designated field and grade level.[3]

---

[1] James B. Conant, The Education of American Teachers (New York: McGraw Hill, 1963), p. 54.

[2] Ibid., p. 55.

[3] Ibid., p. 60.

Conant's proposal suggests that both academic and pedagogical professors, supervisors of student teaching, cooperating public school teachers and administrators, State Department of Public Instruction personnel, and possibly others be brought in to evaluate practice teaching and to judge the candidate's mastery of the subject he teaches, his utilization of educational knowledge, his mastery of the techniques of teaching, and his possession of the intellectual and personality traits relevant to effective teaching.[4] The Conant plan also calls for teacher education institutions, in conjunction with public school systems, to establish a state-approved practice-teaching arrangement, and stipulates that public school systems which enter contracts with teacher education institutions for practice teaching should designate as classroom teachers working with practice teaching only those persons in whose competence as teachers, leaders, and evaluators they have the highest confidence.[5]

This study was proposed to demonstrate and field test Conant's plan. The nationwide implications of this plan for teacher education and certification are clear to those concerned with quality teaching. It deserves a wide-scale demonstration to determine if it can be made workable. It should not be rejected out of hand or be installed as the prototype in the education and certification of teachers without a thorough trial and evaluation.

In reaction to the Conant recommendations a number of problems involved in basing certification on judgments of teacher competence can be pointed out. For example:

1. There is widespread skepticism among professional educators that alternative teacher education programs which rely on use of judgments of classroom teaching quality can be (a) practical, (b) reliable, or (c) acceptable to State Departments of Education as a basis for legal certification of teachers.

2. Related to this skepticism, it is generally assumed that lack of agreement regarding what should be the criteria for "good" teaching constitutes an insurmountable roadblock to basing certification on demonstrated competence.

3. There is also a popular belief that judgments of teaching competence must somehow be "objective," and that rating procedures which involve subjective value judgments are unreliable and dangerous.

4. Established teacher education and certification practices have been strongly influenced by the notion that there is or should be "one best" definition of good teaching and "one best" plan for the preparation of teachers.

---

[4] *Ibid.*, p. 62.

[5] *Ibid.*, p. 63.

Problems of this sort had to be dealt with in order to give the Conant recommendations any kind of a fair trial. Therefore, consideration of them was included in the objectives of the study.

Objectives

The general objectives of this study were:

A. To demonstrate and evaluate the use of judgments of competence in classroom performances as a potential basis for certifying teachers.

B. To make wider use of and test out in practical field situations basic research knowledge of various processes related to judging teacher competence from observations of classroom performance, and, if possible, add to that knowledge.

C. To develop and test procedures by which observer-judges can evaluate teaching behavior using individual criteria identifiable in terms of measured positions on relevant value continua.

D. To involve both academic and pedagogical scholars in an all-university approach to the process of evaluating the qualifications of candidates for teacher certification.

E. To develop working partnerships between teacher education institutions, State Departments of Public Instruction, and local school systems that involve both shared responsibility for teacher education and cooperative judgments of candidates' qualifications for teaching.

F. To provide descriptions of variation and central tendencies in the performance of student teachers.

G. To provide descriptive information about observer-judge ratings of teacher competence, including the identification of factors influencing their reliability and validity, as well as the variation and central tendencies of their observations and evaluations.

Research Foundation

Basic to this demonstration are the recommendations of James B. Conant, which are the product of his recent two-year study of the education of American teachers.[6] Although the Conant study may not qualify as basic research, it does represent a painstaking survey and penetrating assessment of the complex and controversial issues involving the education and certification of teachers conducted by a most respected scholar, scientist, and statesman, who was assisted by a strong staff of qualified educators. Dr. Conant's recommendations are rooted in the findings of a study of current facts and issues, which is considerably more than can be said for the rooted-in-tradition policies and procedures his recommendations are designed to displace.

---

[6] *Ibid*.

With respect to the demonstration and evaluation of procedures for observing classroom performance and judging teaching competence, there is, fortunately, a considerably stronger research foundation. Most pertinent to the implementation of the proposed program is the vast amount of research on teaching drawn together under the editorship of N. L. Gage.[7] There are available a wide variety of experimentally tested procedures for measuring classroom behavior by systematic observation,[8] for rating competence in teaching,[9] for analyzing teaching methods,[10] for analyzing the teacher's personality and characteristics,[11] and for measuring cognitive[12] and noncognitive variables[13] in research on teaching, which will be selected and used in this program.

---

[7] N. L. Gage (ed.), Handbook of Research on Teaching (Chicago: Rand McNally & Co., 1963).

[8] Donald M. Medley and Harold E. Mitzel, "Measuring Classroom Behavior by Systematic Observation," Handbook of Research on Teaching, N. L. Gage (ed.) (Chicago: Rand McNally & Co., 1963), pp. 247-328.

[9] H. H. Remmers, "Rating Methods in Research on Teaching," Handbook of Research on Teaching, N. L. Gage (ed.) (Chicago: Rand McNally & Co., 1963), pp. 329-378.

[10] Norman E. Wallen and Robert M. W. Travers, "Analysis and Investigation of Teaching Methods," Handbook of Research on Teaching, N. L. Gage (ed.) (Chicago: Rand McNally & Co., 1963), pp. 448-505.

[11] J. W. Getzels and D. W. Jackson, "The Teacher's Personality and Characteristics," Handbook of Research on Teaching, N. L. Gage (ed.) (Chicago: Rand McNally & Co., 1963), pp. 506-582.

[12] Benjamin S. Bloom, "Testing Cognitive Ability and Achievement," Handbook of Research on Teaching, N. L. Gage (ed.) (Chicago: Rand McNally & Co., 1963), pp. 378-397.

[13] George G. Stern, "Measuring Noncognitive Variables in Research on Teaching," Handbook of Research on Teaching, N. L. Gage (ed.) (Chicago: Rand McNally & Co., 1963), pp. 398-447.

This investigation, while not designed primarily as a basic research study, was compelled, in making intelligent use of available procedures for measuring and judging teacher behavior, to draw upon and add some increment to the basic research knowledge in this area. The observation, judging, and evaluation phases of the program were conducted with as much research rigor as possible under the circumstances. Likewise, the data collected were submitted to the most strenuous statistical analyses that could be found or developed.

CHAPTER II

RESEARCH PROCEDURES

General Design

The study was carried out in four phases over a period of four years (1964-1968). Phase I involved the selection, organization, and orientation of observer-judges, as well as the selection and evaluation of the observational and rating instruments and procedures. Phase II involved observer-judges making multiple and repeated observations and judgments of student teachers' classroom performances in pre-service clinical experiences. Phase III was a follow-up study involving observations and judgments of a sample from Phase II subjects during their first year of service as certified teachers. Phase IV was concerned with analysis and evaluation of the data, and the preparation of the final report.

Subjects

The prospective teachers observed and judged were drawn from those students enrolled in teacher education programs at Sacramento State College in California, the University of Florida at Gainesville, New York State University at Albany and at Oneonta, Northwestern University in Evanston, Illinois, and the University of Wisconsin in Madison. The teams of observer-judges were selected from the faculty of these colleges and cooperating public school systems under the supervision and approval of the State Department of Public Instruction in each of the five states involved.

Six populations or groups of subjects were involved in this study:

(1) Observer-judges of filmed teaching episodes (Raters_1).

(2) Observer-judges of pre-service teaching performances (Raters_2).

(3) Observer-judges of in-service teaching performances (Raters_3).

(4) Five master teachers recorded on film (Ratees_1).

(5) Pre-service student teachers (Ratees_2).

(6) In-service teachers (Ratees_3).

Raters_1 consisted of college supervisors of student teaching, education professors, and professors of academic subjects from four of the six teacher education institutions, Sacramento, Albany, Northwestern, and Wisconsin.

Raters$_2$ consisted of student teacher supervisors, education professors, and academic professors from all six colleges, plus cooperating teachers and principals from the public schools and, in a few cases, personnel from State Departments of Public Instruction.

Raters$_3$ included not only members of the raters$_2$ group but also teachers, supervisors, and administrators from the schools where the ratees$_2$ were undergoing their initial in-service experience. Although membership of the rater groups varied from phase to phase there were individual observer-judges who participated in all three rater groups.

Ratees$_1$ consisted of five master teachers whose teaching had been recorded on film. These teachers were members of the faculty of Wisconsin High School (a private school operated by the University of Wisconsin until 1962) in 1959 and 1960 when the films were made. They were no longer available for study beyond their performances as teachers on the films.

Ratees$_2$ was comprised of a sample of about 500 elementary and secondary student teachers from the six teacher education institutions. This was the group observed and evaluated in Phase II.

Ratees$_3$ was comprised of 100 first-year teachers selected from the Ratees$_2$ group, plus 100 experienced teachers who were selected from the schools in which the first-year teachers were employed.

Instrumentation

Two sets of instruments were used in the collection of data for this study: the first set, the *Study of Beliefs*, was used to assess the beliefs of all raters and ratees; the second set, the *Teacher's Classroom Behavior*, was used by the raters to observe, record, and evaluate the teaching performances of the ratees. The *Study of Beliefs* was comprised of three instruments: the *Personal Beliefs Inventory* (PBI), the *Teacher Practices Inventory* (TPI), and the *Dogmatism Scale* (D-Scale). The *Teacher's Classroom Behavior* included three instruments: the *Teacher Practices Observation Record* (TPOR), the *Classroom Behavior Rating Scale* (CBRS), and the *Teacher Evaluation Scale* (TES). Each of these instruments is described as follows:

Personal Beliefs Inventory. The Personal Beliefs Inventory (PBI) is a yardstick by which agreement-disagreement with the basic philosophy of John Dewey may be measured. A high score on this inventory indicates that one's beliefs concur with Dewey's fundamental philosophic beliefs. Reliabilities reported for the PBI vary from .55 to .78.[1]

---

[1] Bob Burton Brown, *The Experimental Mind in Education* (New York: Harper and Row, 1968), Chapter VI.

Teacher Practices Inventory. The Teacher Practices Inventory (TPI) measures agreement-disagreement with Dewey's educational philosophy. A high score indicates concurrence with Dewey's beliefs about what teachers should do in classrooms. Reliability coefficients reported for the TPI range from .56 to .94.[2]

Dogmatism Scale. The Dogmatism Scale (D-Scale) measures the structure of belief systems along an open and closed dimension. Reliabilities of .68 to .93 have been reported for the D-Scale.[3]

Teacher Practices Observation Record. The Teacher Practices Observation Record (TPOR) is a sign system for recording teacher practices observed in a classroom. It measures the agreement-disagreement of teachers' observed classroom behavior with educational practices advocated by Dewey in his philosophy of experimentalism. A high score on the TPOR indicates that the recorded behavior was observed as practices which Dewey advocated. Establishment of reliability of the TPOR is reported in Phase I of the study.

Classroom Behavior Rating Scale. The Classroom Behavior Rating Scale (CBRS) is a scale on which descriptive dimensions of teacher and pupil behavior are rated on a six-point continuum. This scale was developed from rating instruments used by Ryans and McGee in earlier studies.[4]

Teacher Evaluation Scale. The Teacher Evaluation Scale (TES) is a six-point scale (enlarged in Phase III to an eighteen-point scale in order to permit observers to make finer discriminations). The TES is an instrument which enables the rater to evaluate the competence of the teacher observed.

These instruments may be found in Appendix A and B.

Data Collection

Phase I. In the spring of 1965, viewing sessions of filmed episodes of teaching behavior were held at four of the participating teacher training institutions. During these sessions observer-judges (raters$_1$) were acquainted with the use of the TPOR, making recorded observations of each of five filmed episodes. Twelve months later

---

[2] Ibid.

[3] Milton Rokeach, The Open and Closed Mind (New York: Basic Books, Inc., 1960).

[4] David G. Ryans, Characteristics of Teachers (Washington, D.C.: The American Council on Education, 1960).

H. M. McGee, "Measurement of Authoritarianism and Its Relation to Teacher's Classroom Behavior," Genetic Psychology Monograph, 52:89-146, 1955.

raters_1 viewed two of these films, repeating the Teacher's Classroom Behavior observations. Prior to the initial viewing sessions all observer-judges completed the Study of Beliefs.

Phase II. To the rater_1 group additional personnel (cooperating teachers and administrators) were added for Phase II; these observer-judges comprised rater_2 group. The ratees_2 group was composed of student teachers completing their pre-service clinical teaching experiences in the spring of 1966. Both the groups of observer-judges and student teachers completed the Study of Beliefs. During the student teaching experience each pre-service teacher's classroom behavior was repeatedly observed by teams of observer-judges using the Teacher's Classroom Behavior as an observation and rating instrument.

Phase III. Additional ratees (experienced teachers) and raters (public school personnel) were added as subjects in Phase III. The Study of Beliefs was completed by each of the added personnel. Each of the ratees_3 was observed by a team of three observer-judges systematically during the winter and spring of 1967.

Evaluation of Data

Phase I. The scores from the Study of Beliefs and Teacher's Classroom Behavior were analyzed to:

(1) develop a design for estimating reliability coefficients for the recorded observations of the observer-judges

(2) identify variables that could be used in predicting the observation scores and ratings given a teacher by an observer-judge.

Phase II. Phase II data were analyzed to identify:

(1) the relationships between the beliefs and observed practices of the ratees_2

(2) the relationships between beliefs, observations, and evaluations of the raters_2

(3) the interaction of these relationships

(4) variables which contribute information to the prediction of ratings given teachers by observer-judges.

Phase III. Phase III data were analyzed to:

(1) identify the relationships between the beliefs and observed practices and evaluations of ratees_3

(2) identify the relationships between the beliefs, observations and evaluations of raters_3

(3) identify the interactions between these relationships

(4) compare the raters$_2$' and raters$_3$' observation scores and evaluation scores of pre-service teachers (ratees$_2$) to those of the same individuals as first-year teachers (ratees$_3$)

(5) compare the observations and evaluations of first-year teachers to those of experienced teachers given by raters$_3$

(6) identify characteristics of raters which become predictive of ratings given certain characteristics and behavior of ratees

(7) identify characteristics and behavior of ratees which become predictive of ratings of teacher competence

(8) identify the relationships of ratings in Phase II with ratings in Phase III for (a) all ratees (b) ratees given extremely good ratings compared with ratees given extremely poor ratings.

CHAPTER III

PHASE I - PILOT STUDY

Purposes

The objectives of Phase I were (1) to select and orient the observer-judges who were to serve in the study, (2) to acquaint these observer-judges with the observation and rating instruments, (3) to measure their relevant value positions, (4) to establish estimates of their reliability as classroom observers, and (5) to identify variables that could be used in predicting the observation scores and ratings given a teacher by the observer-judges.

Subjects

Subjects - Observer-Judges. The observer-judges, all volunteers, were student-teaching supervisors, education professors, and professors of academic subjects drawn from the faculties of four of the teacher training institutions[1] that participated in the study. These observer-judges (raters$_1$) recorded observations of filmed teaching episodes during the spring of 1964 and again a year later in the spring of 1965. A total of 130 subjects served as observer-judges for this phase.

Subjects - Ratees. The ratees were five experienced teachers whose teaching behavior had been filmed at the University of Wisconsin. These teachers all held master's degrees and had been selected as outstanding teachers. From the unedited films, 50 to 60 minutes in length, 30-minute continuous segments were cut. Selection of the films and of the segments taken from them was made for purposes of achieving variety in teaching style, grade level, and subject matter taught. Film #1 was of a ninth-grade French class; Film #2, a seventh-grade mathematics class; Film #3, a fourth-grade unit on "Weather"; Film #4, a ninth-grade speech class; and Film #5, a seventh-grade science class. These five films were used (1) for the orientation of the observer-judges, (2) to gather data for reliability studies, and (3) to gather data for identification of variables for predicting observation and rating scores.

Data Collection Procedures

Prior to the viewing sessions the observer-judges completed the Study of Beliefs; for each observer-judge data were gathered which gave a quantitative score for the measurement of personal and educational beliefs. In addition, information regarding the sex, age, occupation, and institutional affiliation of each subject was obtained.

---

[1] University of Wisconsin, Northwestern University, Sacramento State College, and State University of New York at Albany.

During a six-weeks interval, film observation and recording sessions were held at each of the campuses of the four participating institutions. Conditions of the viewing sessions were similar. All observer-judges received the same 10-minute explanation, by the same person, for recording both their observations in the Teacher Practices Observation Record and their ratings on the Classroom Behavior Rating Scale and Teacher Evaluation Scale. During the viewing of Film #1, time was called in order for the observer-judges to become familiar with the observational procedures and instrumentation. This constituted the orientation provided the observers. As one of the major purposes of the study was to investigate observation and rating of teachers on the basis of a rater's individual criteria, no attempt was made to bring the observer-judges to agreement concerning their recorded observations, nor was any discussion to this effect permitted. For the other four films, no assistance of any kind was given the observer-judges. Each observer-judge, for each film viewed, completed the set of observation and rating instruments.

Mass observations of films are expensive and administratively difficult to arrange. For these reasons, repeated observations the second year could be obtained on only two of the five films. Film #1 was eliminated because it had been used as the orientation film and conditions of the first viewing could not be duplicated. Data obtained from the first viewing of Film #3 indicated a wide discrepancy in scores based on viewing locations, which could have been due to the artificial conditions under which it was filmed. Film #5 had not been observed at all four institutions. This left Films #2 and #4 for the second viewing. It was possible to obtain repeated TPOR scores on these two films from only a portion of those who observed the first viewings.

Reliability

Reliability Estimates
Teacher Practices Observation Record

Analysis of Data

Means and standard deviations of observation scores were computed for each film for each viewing session. Means were examined to determine if significant differences in TPOR scores were given at the four participating institutions or by the three major occupational classifications of observer-judges.

Data were submitted to analysis of variance to develop a between-observer reliability coefficient. The data were also used to develop statistical procedures for establishing within-observer reliability estimates. Lastly, the data were submitted to the Kuder-Richardson Formula 20 for measuring item reliability.

Findings - TPOR Scores

Mean Scores. Table 1 shows the mean TPOR score given each of the five films by the observer-judges on the first viewing. The French teacher in Film #1 was seen as the least experimental and the fourth-grade teacher in Film #3 as the most in agreement with Dewey. The range of more than 40 points between the high and low TPOR means indicates the ability of the instrument to differentiate various styles of teaching.

Table 1

Mean TPOR Scores Given Five Films
by All Observers

Film	No. of Observers	Mean	S.D.
No. 1	130	80.01	13.32
No. 2	124	115.86	16.84
No. 3	119	120.96	22.74
No. 4	119	104.24	17.10
No. 5	67	98.84	12.88

Differences in the mean TPOR scores given at the four different participating institutions were examined. The location variable was found to have little or no influence. Using Scheffé's comparisons, no statistically significant differences were found among the TPOR means given at the various locations for Films #1, #2, #4, and #5. The only statistically significant differences were found between California and each of the other three locations on Film #3.

Differences in the mean TPOR scores given by the three major occupational classifications of observer-judges--college supervisors of student teaching, education professors, and academic professors--were also examined. No statistically significant differences were found between any of these groups for Films #1, #2, #4, and #5. The only statistically significant differences were found between supervisors of student teaching and both education and academic professors on Film #3.

Scores for the two viewings of Films #2 and #4 were compared and are reported in Table 2. This shows a fairly substantial difference between TPOR means recorded for the first and second viewings of Film #2. While this difference raises some questions about stability, both means for this subgroup of 69 observers lie well within one standard deviation of the mean of 115.86 for 119 first-viewing observers which may simply demonstrate the normal variability of TPOR scores. The differences between TPOR scores for the first and second viewings of Film #4 are very small.

Table 2

Mean TPOR Scores Given Films On
Repeated Observations One Year Apart

Film	Viewing	No. of Observers	Mean	S.D.
No. 2	1st	69	122.22	20.52
No. 2	2nd	69	109.81	18.31
No. 4	1st	72	107.15	17.15
No. 4	2nd	72	105.14	18.12

Reliability Coefficients. Reliability of instruments of measurement is a complex concept which becomes compounded when dealing with the measurement of classroom behavior by systematic observation. The question of the reliability of the observers and the recording of their observations must be added to the problem of instrument reliability. In the past most observational studies have limited their study of observer reliability to computing the correlation between two sets of observations or to figuring the percent of agreement between observers. Following this procedure, the correlations between the TPOR scores obtained from the repeated observations of Films #2 and #4 were computed and are reported in Table 3. The correlations of the columns (10-minute observation periods) within each film observation are very high, but the correlations between the 1964 and 1965 observations are very low. The first indicates that the observers tended to maintain the same relative position in the group throughout the viewing of a single film on a given day. The second indicates that sizable shifts in these positions took place during the intervening year. In other words there was good consistency within one occasion or viewing, and again within another, but poor stability between two widely separated occasions. One must keep in mind, however, that such reliability coefficients normally decline proportionately with the length of time between "tests." Had the repeat observations been made only a month or so apart considerably higher correlations might have been expected.

Even so, correlation of two sets of scores by a number of different observers is not likely to be a very accurate estimate of reliability. It is difficult to make arrangements for large numbers of observers to view the same classroom on two different occasions, or to control variations between those occasions. Likewise, the number of

Table 3

Correlation of TPOR Scores Obtained from Repeated Observations of Films

FILM NO. 2

TPOR Column		1964 Observation				1965 Observation			
		TPOR Column				TPOR Column			
		1	2	3	TOT	1	2	3	TOT
1964 Observation	1	1.00	.79	.69	.89	.36	.25	.12	.27
	2	--	1.00	.81	.95	--	.29	.16	.31
	3	--	--	1.00	.92	--	--	.20	.29
	TOT	--	--	--	1.00	--	--	--	.32
1965 Observation	1	--	--	--	--	1.00	.61	.55	.80
	2	--	--	--	--	--	1.00	.81	.93
	3	--	--	--	--	--	--	1.00	.90
	TOT	--	--	--	--	--	--	--	1.00

FILM NO. 4

TPOR Column		1964 Observation				1965 Observation			
		TPOR Column				TPOR Column			
		1	2	3	TOT	1	2	3	TOT
1964 Observation	1	1.00	.75	.52	.86	.32	.36	.25	.34
	2	--	1.00	.71	.93	--	.46	.52	.52
	3	--	--	1.00	.85	--	--	.67	.65
	TOT	--	--	--	1.00	--	--	--	.57
1965 Observation	1	--	--	--	--	1.00	.79	.71	.90
	2	--	--	--	--	--	1.00	.83	.95
	3	--	--	--	--	--	--	1.00	.92
	TOT	--	--	--	--	--	--	--	1.00

classrooms observed on two different occasions by two different observers is likely to be small. In either case, the size of the N determines the precision of the correlation coefficient, and since the N of even well-financed observational studies rarely exceeds 100 the confidence intervals for the coefficients are extremely wide. Furthermore, such correlations are usually based on total scores which ignore variations in scoring individual items or categories. It is possible to obtain a perfect correlation of total scores when the reliability for the items is zero. If on a 70-item "sign" system, for example, the 35 odd-numbered items are marked "+" and the 35 even-numbered items are marked "0" on the first observation, and then exactly reversed on the second observation, identical total scores will be obtained and used to produce a deceivingly perfect reliability correlation.

Percent of agreement between observers tells almost nothing about the accuracy of the scores obtained. It is entirely possible to find observers agreeing 99 percent in recording behaviors on an instrument whose item or category consistency is very poor. Reliability can be low even though observer agreement is high for several reasons. For example, observers might be able to agree perfectly that a particular teaching practice occurred in a classroom, yet if that same practice occurs equally, or nearly so, in all classrooms, the reliability of that item as a measure of differences between teachers will be zero. Near-perfect agreement could also be reached about the percentage of time a number of teachers employed certain categories of behavior; but if every teacher sharply reversed these percentages from period-to-period or day-to-day, the reliability of these categories would be zero. Errors arising from variations in behavior from one situation or occasion to another can far outweigh errors arising from failure of two observers to agree exactly in their records of the same behavior.

Yet, the reliability of most instruments for systematically recording the behavior of teachers requires a high percent of observer agreement. "Between-observer" agreement has become almost a cardinal principle in planning observational studies. According to Medley and Mitzel a sample of classrooms from the population to be studied should be visited by trained recorders using the observational instrument in the same way it will be used in any subsequent study. In order to study the "objectivity" of the items, i.e., how closely observers agree in recording identical behaviors, at least two recorders should be present on each visit, sitting in different parts of the room and making independent records. In order to be able to estimate how stable the two records based on different visits will agree, each class should be visited at least twice. To recapitulate, in their words, "_c_ teachers are visited in _s_ situations by a team of _r_ recorders. In studying the reliability of a scale with _i_ items on it, the total number of scores to be analyzed will be _cris_."[2]

To match this rigorous plan for data collection Medley and Mitzel have taken the classic definition of reliability, $\rho_{xx} = \frac{\sigma_{\tau}^2}{\sigma_x^2}$ and applied it to measurements of classroom behavior. In this definition, true variation, σ_{τ}^2 is defined to be the variation of the total score for any class (teacher) when the effects of recorders (observers), items on the scoring instrument, and situations (viewings or visits) have been removed. The true variation plus "error," σ_x^2 is defined to be the variation of the total scores for any class, including variation contributed by items on the scoring instrument, recorders, situations and random error. The smaller the effect of the recorders, items, and situations for a class total, the higher the reliability coefficient will be. In other words, if the instrument has high reliability, the scoring of the class or teacher is relatively free of the effects of recorders, items, or the different situations under which the scoring was done, and as such, reflects a "good" or reliable instrument.[3]

---

[2] Donald M. Medley and Harold E. Mitzel, "Measuring Classroom Behavior by Systematic Observation," _Handbook of Research on Teaching_, N. L. Gage (ed.) (Chicago: Rand McNally & Co., 1963), p. 309.

In seeking a design for estimating the reliability of TPOR observations, we closely examined the four-way analysis of variance model suggested by Medley and Mitzel. While it was found to be a sound approach to reliability estimation, it may not be entirely appropriate for analyzing the data obtained in the film study described above. For instance, in the simple example given by Medley and Mitzel in the Handbook of Research on Teaching, page 316, where one item is used to score 24 classes (teachers) observed during four situations by two recorders (observers), the reliability coefficient is estimated by:

$$\rho_{xx} = 1 - \frac{MS_{cxr}}{MS_c}$$

where MS_{cxr} is the mean square for classes x recorders obtained from the analysis of variance table and MS_c is the mean square for classes obtained from the analysis of variance table. The coefficient of reliability in this case actually reflects not instrument reliability, but rather, recorder or observer reliability. When MS_{cxr} is large, it indicates an inconsistency on the part of the observers to score the classes in the same way, which in turn causes ρ_{xx} to be small. In like manner, a very small value of MS_{cxr} reflects consistency in scoring, in which case ρ_{xx} will be large.

Training of the observers undoubtedly would bring them into agreement with respect to recording or scoring identical behaviors, which would be reflected in a higher reliability coefficient, ρ_{xx}. However, in the previously described film study in which the TPOR was tried out, no attempt was made to train the observers. To the contrary, a deliberate attempt was made to preserve the differences among observers by selecting them from varying occupational groups, from varying sizes of institutions with varying orientations to teacher education, and from varying parts of the country. We wanted to test the reliability of the TPOR under uncontrolled field conditions to see what value it might have in the hands of the differing kinds of people who carry out the everyday responsibilities for teacher education in America. Hence, the component of variance due to the observers' variability in this study would cause σ_x^2 to be large compared to σ_τ^2, resulting in a small ρ_{xx}. There was not as much observer variability as might have been expected, however. When the Medley-Mitzel model was adapted to fit the film study data the TPOR observations were found to have a modest but substantial reliability coefficient of .57.

In the analysis of variance example cited above it should also be noted that two of the variables of interest, viz., classes and situations, had but one degree of freedom each. This being the case, "poor" estimates of the components of variance could result. In fact, the components of variance could be estimated to be zero (which happens in many cases). Also, since the estimate of ρ_{xx} would consist of the ratio of linear combinations of mean squares, the bounds of error on this estimate could be exceedingly large.

---

[3] Ibid.

The unsuitability of the Medley-Mitzel model for this data results primarily, however, from the fact that it stresses "between-observer" variability rather than "within-observer" variability. This is a philosophical rather than a statistical issue. Reliability coefficients which reward high agreement between observers imply that one should seek a single, uniform, "objective" system for observing and classifying teaching behavior. From the point of view of the framework underlying the development of the TPOR, objectivity in perceiving and quantifying such behavior is neither possible nor desirable. "Between-observer" agreement may not only encourage a false sense of confidence with respect to the accuracy of measurements, but also gives a false sense of "objectivity" regarding the observations. A team of observers can be trained to the point of near-perfect agreement, but this does not erase the possibility that instead of several differing "subjective" judgments, they now make only one. Therefore, another mathematical definition of reliability was sought, one which is concerned primarily with "within-observer" variability.

It was reasoned that if having scored a given filmed teaching situation, the same observer-judge were to score the same teaching situation again in the same way, then it could be said the observer-judge's scoring was reliable. Hence, a definition for "within-observer" reliability for a given observer-judge and film was devised as follows:

	Viewing		
Items	1	2	$d_i = x_{1i} - x_{2i}$
1	x_{11}	x_{21}	d_1
2	x_{12}	x_{22}	d_2
3	x_{13}	x_{23}	d_3
.	.	.	.
.	.	.	.
.	.	.	.
n	x_{1n}	x_{2n}	d_n

Consider the variances of the differences d_i, where

$$d_i = x_{1i} - x_{2i}.$$

If the scores are independent, i.e., the judge is not consistent, or in fact marks by chance, then

$$
\begin{aligned}
V(d_i) &= V(x_{1i} - x_{2i}) \\
&= V(x_{1i} + V(x_{2i}) \\
&= \sigma^2 + \sigma^2 \\
&= 2\sigma^2 \quad \text{(or 2 Var(x))}
\end{aligned}
$$

However, if the judge is consistent from viewing to viewing, his 2 scores should be positively correlated and now

$$
\begin{aligned}
V(d_i) &= V(x_{1i} - x_{2i}) \\
&= V(x_{1i}) + V(x_{2i}) - 2\ \mathrm{Cov}\ (x_{1i}, x_{2i}) \\
&= 2\sigma^2 - 2\sigma_{12}
\end{aligned}
$$

or

$$V(d_i) = \sigma_d^2 = 2\sigma^2 - 2\sigma_{12}$$

It is noted that the following assumptions are made in the above discussion:

1) The variance of each item score is the same for all items over viewings; i.e.,

$$V(x_{ij}) = \sigma^2 \quad \text{for} \quad \begin{aligned} i &= 1,2 \\ j &= 1 \ldots n \end{aligned}$$

2) Under the complete randomness assumed under chance scoring, each value of x is assumed to have equal chance of being selected; hence

$$p(x) = \frac{1}{k}$$

where k is the number of choices available.

Now we define for judge j and film f,

$$\rho_{jf} = 1 - \frac{\sigma_d{}^2}{2\sigma^2}$$

where

$$
\begin{aligned}
\sigma_d{}^2 &= \mathrm{Var}\ (d_i) && i = 1 \ldots n \\
\sigma^2 &= \mathrm{Var}\ (x_{ij}) && i = 1,2 \\
& && j = 1 \ldots n
\end{aligned}
$$

However, under the assumptions of a random choice by the judge, σ^2 becomes a constant, computed as

$$\sigma^2 = \sum_x (x - \mu)^2\ p(x)$$

We calculate the sample value of s_d^2 and use it to estimate σ_d^2.

Hence we are working with a statistic

$$r_{jf} = 1 - \frac{s_d^2}{2\sigma^2}$$

Now, if there is in fact high positive correlation of the scoring from viewing 1 to viewing 2, then

s_d^2 will be small (i.e., s_{12} will be large)

and

r_{jf} will be close to 1.

If the scoring from viewing to viewing is in fact independent and really associated with a chance event, then

s_d^2 will be of the magnitude of $2\sigma^2$ (i.e., s_{12} will be small; close to zero)

and r_{jf} will be close to 0.

The coefficient r_{jf} will theoretically be in the interval (0,1) where a maximum value of one implied absolute correlation, while a minimum value of zero implies the same scoring could have happened by chance, hence no reliability. However, the possibility of $r_{jf}<0$ exists because there is a non-zero probability that the scorings will be negatively correlated and this may cause s_d^2 to be greater than σ^2; this in turn causing $r_{jf} <0$.

Worth mentioning is the fact that this statistic uses a larger than expected variance σ^2, as a yardstick against which the judge's variation from viewing to viewing is compared. This is because one would expect a judge to select the extremes in scoring an item less frequently than scores near the center of the scale; such scoring would likely yield a variance smaller than that implied by a completely random selection. This yardstick could, in effect, cause the coefficient r_{jf} to be depressed as compared with other measures of reliability.

Using the above formulation the "within-observer" reliability of TPOR scores was computed for the two filmed teaching situations on which repeated viewings were made a year apart. Table 4 shows eight reliability coefficients ranging between .48 and .62.

These coefficients of reliability reflect observer reliability rather than instrument reliability. In order to determine the internal consistency of the TPOR, its item reliability, which would indicate something of its potential in the hands of reliable observers, the film study data were submitted to Kuder-Richardson Formula 20 for estimating item reliability. Table 5 shows these results.

Table 4

"Within-Observer" Reliability Coefficients for TPOR Scores on Repeated Viewings of Films

	FILM NO. 2	
	N = 69	
TPOR Column	r_{jf}	error
TOT	.48	.0255
1	.57	.0177
2	.51	.0194
3	.51	.0177
	FILM NO. 4	
	N = 72	
TPOR Column	r_{jf}	error
TOT	.52	.0191
1	.56	.0182
2	.57	.0244
3	.62	.0171

Table 5

TPOR Internal Consistency Reliability Coefficients

			TPOR Columns			
Film	Viewing	N	1	2	3	TOT
No. 1	1st	158	---	---	---	.86
No. 2	1st	69	.79	.81	.83	.93
No. 2	2nd	69	.77	.81	.79	.91
No. 3	1st	140	---	---	---	.93
No. 4	1st	72	.76	.77	.78	.90
No. 4	2nd	72	.76	.78	.77	.91
No. 5	1st	84	---	---	---	.85

If each item is highly correlated with every other item on the instrument, then the instrument has good item reliability or internal consistency. The fact that the TPOR scores yielded uniformly high internal reliability coefficients is not surprising in light of the fact that throughout their development the TPOR, TPI, and PBI underwent repeated RAVE analysis, an iterative procedure which yields a set of item response weights which maximize the internal consistency of inventories.[4]

TPOR Reliability Summary

Having submitted this instrument to the hazards of uncontrolled use by uncontrolled observers, and then submitting it to the severest statistical procedures that could be found, we can make the following conclusions as to reliability estimates for the Teacher Practices Observation Record: (1) correlation of observers' total scores within a given film viewing--VERY GOOD, (2) correlation of observers' total scores between repeat film viewings one year apart--POOR to FAIR, (3) between-observer reliability--FAIR, (4) within-observer reliability--FAIR, (5) internal consistency reliability--VERY GOOD.

Reliability Estimates
Classroom Behavior Rating Scale

The Classroom Behavior Rating Scale is an instrument used in this study to rate the behavioral characteristics of teachers and their students on an authoritarian-egalitarian dimension.[5] The scale consists of thirteen items which describe teacher characteristics, Classroom Behavior Rating Scale-Teacher (CBRST) and four items which describe pupil characteristics, Classroom Behavior Rating Scale-Pupil (CBRSP). Each item is scored by the observer on a six-point continuum; the higher the score, the more authoritarian the behavior observed. A maximum score of 102 indicates extreme authoritarian behavior, and a minimum score of 17 indicates non-authoritarian behavior.

Analysis of Data

Means and standard deviations of rating scores were computed for each film. Means were examined to determine if significant differences in CBRS scores were given at the four participating institutions or by the three major occupational classifications of observer-judges.

---

[4] Ronald Ragsdale and Frank B. Baker, The Method of Reciprocal Averages for Scaling of Inventories and Questionnaires: A Computer Program for the CDC 1604 Computer, (Mimeographed, Laboratory of Experimental Design, Department of Educational Psychology, U. of Wis., Madison).

[5] Description of instrument can be found in Chapter II, p. 8.

Through the use of the same procedures developed for estimating within-observer reliability of the Teacher Practices Observation Record (reported in the previous section of this report) within-observer reliability coefficients were developed. Data were also submitted to the Kuder-Richardson Formula 20 to establish internal-consistency measures for the instrument.

Findings - CBRS Scores

Mean Scores. Table 6 shows the mean Classroom Behavior Rating Scale scores given each of the five films by the observer-judges on the first viewing.

Table 6

Mean CBRS Scores for Films

Film	No. of Observers	Mean	S.D.
No. 1	130	47.80	11.68
No. 2	124	37.21	9.94
No. 3	119	35.94	11.64
No. 4	119	41.70	14.36
No. 5	67	42.54	11.73

The French teacher in Film #1 received the highest score, hence was seen as the most authoritarian, while the fourth-grade teacher in Film #3 was seen as the least authoritarian by virtue of receiving the lowest mean score. The range of slightly less than twelve points between the high and low CBRS means indicates the rather limited ability of the instrument to differentiate between teachers.

The CBRS scores given at the four participating institutions were examined for differences and it was determined that the location variable had little or no influence on mean scores. No statistically significant differences were found among the CBRS means given at the various locations in Films 1, 2, 4, and 5. The only differences having statistical significance were between California and each of the other three locations on Film #3.

CBRS scores were also examined for differences with respect to the three major occupational classifications of the observer-judges--clinical supervisors, education professors, and academic professors. Again, no statistically significant differences were found for Films 1, 2, 4, and 5 between groups of observer-judges. In Film #3 significant differences were found between the clinical supervisors and both academic and education professors. These differences in mean CBRS scores show a similarity to the statistically significant differences found for the TPOR scores.

Reliability Coefficients. The Classroom Behavior Rating Scale is an adaptation of an instrument developed by Ryans and McGee,[6] who have reported between-observer reliability coefficients for trained observers. As the purpose of this study was to use untrained observer-judges who would be participating in the study over periods of twelve to thirty-six months, and who were deliberately prevented from developing criteria which would enable them to increase their agreement, it was determined that within-observer reliability coefficients rather than between-observer coefficients would be a much more reasonable reliability estimate for this instrument. As observer-judges would be recording behavior over a relatively long period of time and their perceptual differences in observing and rating teacher behavior were encouraged rather than "trained out," the consistency of an observer-judge's ratings over time seemed to be of most importance in establishing observer reliability.

Table 7 reports the within-observer reliability coefficients computed for the two films which had been observed a year apart. These coefficients were determined for the two sections of the instrument, teacher characteristics (CBRST) and pupil characteristics (CBRSP).

Table 7

"Within-Observer" Reliability Coefficients for
CBRS Scores on Repeated Viewings of Films

	FILM NO. 2	
	N = 69	
	r_{if}	error
CBRST	.86	.0191
CBRSP	.84	.0067
	FILM NO. 4	
	N = 72	
	r_{jf}	error
CBRST	.79	.0562
CBRSP	.83	.0208

---

[6] Ryans and McGee (for reference see page 8 of this report).

Within-observer reliability coefficients for the Classroom Behavior Rating Scale range from .79 to .86.

In addition to consideration of observer reliabilty, the internal consistency of the CBRS was examined. As some of the items of the CBRS had been altered or changed in adapting it for use in this study, an analysis of its item reliability was deemed important in order that confidence could be placed in its use. Table 8 shows the results of the CBRS data when submitted to the Kuder-Richardson Formula 20 for estimating internal consistency.

Table 8

CBRST and CBRSP Internal Consistency
Reliability Coefficients

Film	Viewing	N	CBRST	CBRSP
No. 1	1st	130	.79	.48
No. 2	1st	124	.71	.46
No. 2	2nd	69	.76	.57
No. 3	1st	119	.77	.30
No. 4	1st	179	.85	.51
No. 4	2nd	72	.82	.46
No. 5	1st	67	.82	.36

The item reliability for the CBRST ranges from .71 to .85; item reliability for the CBRSP from .30 to .57.

CBRS Reliability Summary

The Classroom Behavior Rating Scale does not indicate from the film data that it is a strong discriminator between behavior of teachers. As it consists of only seventeen items, each of which is scored on a six-point continuum by the observer at the end of a thirty-minute observation period, it tends to measure general impressions of the teacher by the observer rather than discrete teaching behaviors. For this very reason it tends to enjoy good within-observer reliability. The observer would seem to respond to the same teaching behavior as a general perceptual set in much the same manner over a period of time. The CBRST also has good item reliability; the CBRSP item reliability coefficient is, of course, influenced by the very small number of items which comprise this section of the instrument. However, from the film study data we can conclude that the CBRS enjoys good within-observer reliability and adequate internal consistency.

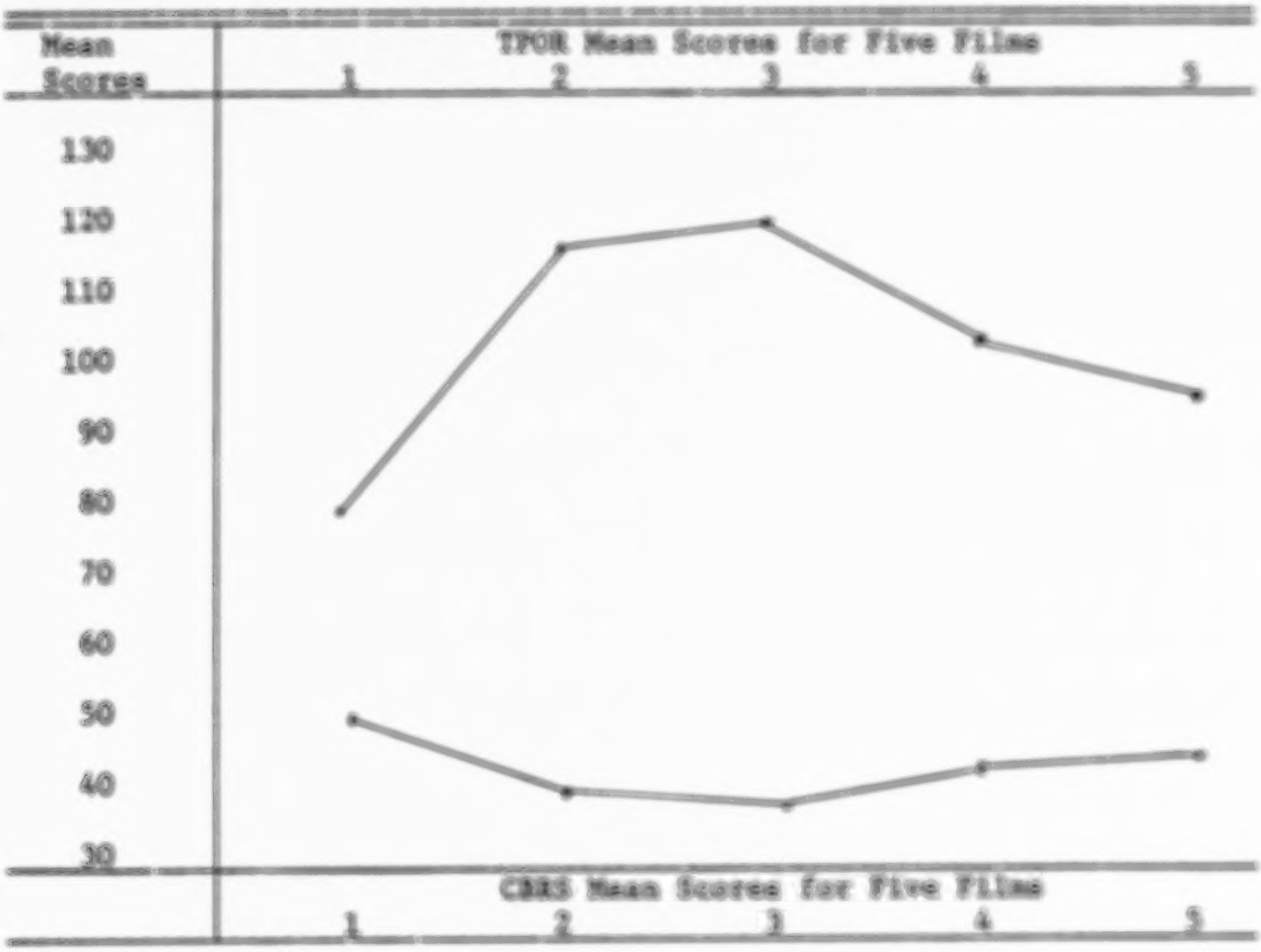
Comparison of the Teacher Practices Observation Record and The Classroom Behavior Rating Scale

The Teacher Practices Observation Record is an instrument for systematically describing the classroom behavior in terms of agreement-disagreement with John Dewey's Experimentalism. The Classroom Behavior Rating Scale is an instrument used to rate the behavioral characteristics of teachers on an authoritarian-egalitarian dimension. From the data of the film study some comparisons can be made of the two instruments.

Figure 1 shows the relationship of the TPOR and CBRS mean scores for the same filmed teaching episodes.

Figure 1

Relationship of TPOR and CBRS Mean Scores



Data from the five film viewings show the inverse consistency in movement and direction of the mean scores on the two instruments. A high TPOR score (agreement with experimentalism) reflects a low CBRS rating (nonauthoritarian behavior characteristics). Thus there seems to be a relationship between experimental behavior and nonauthoritarian behavior characteristics of a teacher. The observer-judges tended to see teachers who are in agreement with experimentalism as nonauthoritarian in behavior and vice versa.

When TPOR means are examined in relation to the evaluative judgments (Teacher Evaluation Scale scores) made about the quality of teaching observed in the films by the observer-judges, an interesting pattern of relationship between TPOR scores and evaluations appears. Table 9 reports these data. While this could mean that the TPOR scores were influenced by how much the observer liked what he saw, the converse is more likely true. The wide differences in TPOR means within each of the evaluative categories are evidence that the relationship between TPOR scores and ratings is relative within the limits describing individual film. In this study, a given TPOR score did not guarantee a "good" or "bad" rating, even though in every case the higher the rating, the higher the TPOR mean score.

Table 9

The Relationship Between TPOR Means and Evaluative Ratings of Five Filmed Teaching Episodes

	Evaluative Ratings					
	A	B	C	D	E	F
Film	Out-standing	Very Good	Good	Fair	Poor	In-competent
No. 1	88.64 (11)	82.21 (48)	79.45 (33)	67.89 (9)	--- (0)	--- (0)
No. 2	126.47 (19)	118.57 (56)	109.19 (21)	--- (0)	--- (0)	--- (0)
No. 3	138.19 (27)	119.32 (38)	109.91 (23)	85.00 (1)	--- (0)	--- (0)
No. 4	110.69 (29)	106.86 (43)	93.73 (11)	76.50 (2)	75.50 (2)	--- (0)
No. 5	115.56 (9)	103.39 (13)	96.00 (23)	88.67 (6)	65.00 (1)	--- (0)

Statistically significant differences beyond the .05 level (using Scheffé's comparison procedures) were found for the following pairs of means:

Evaluative Category A: Films 1 and 2, 1 and 3, 3 and 4, (1 and 4, 1 and 5 were very close)
Evaluative Category B: Films 1 and 2, 1 and 3, 1 and 4, 1 and 5
Evaluative Category C: Films 1 and 2, 1 and 3 (1 and 5 were close)
Film No. 3: Category A and B, A and C

Table 10 shows the correlations between CBRS scores and evaluations given each film. The relationship between CBRS scores and evaluations is relative within limits describing each individual film as evidenced by the rather wide differences in CBRS means within each of the evaluative categories. However, the higher the evaluation, the lower the CBRS mean score; the more authoritarian the teacher is seen, the lower his rating.

Table 10

CBRS Means and Evaluative Ratings

	A	B	C	D	E	F
Film	Out-standing	Very Good	Good	Fair	Poor	In-competent
No. 1	38.83 (11)	43.00 (48)	53.43 (33)	67.11 (9)	0.0 (0)	0.0 (0)
No. 2	29.47 (19)	36.96 (56)	43.90 (21)	0.0 (0)	0.0 (0)	0.0 (0)
No. 3	26.74 (27)	35.87 (38)	46.39 (23)	57.00 (1)	0.0 (0)	0.0 (0)
No. 4	32.21 (29)	41.81 (43)	54.45 (11)	67.50 (2)	70.00 (2)	0.0 (0)
No. 5	24.67 (9)	36.46 (13)	48.22 (23)	59.83 (6)	68.00 (1)	0.0 (0)

Thus, for both the TPOR and CBRS there is a direct relationship between mean scores and evaluative ratings. Within each film the more the teacher is seen in agreement with experimentalism and nonauthoritarianism in behavior, the higher the evaluative rating he receives.

Identification of Predictor Variables

Teacher Practices Observation Record

Analysis of Data. One of the basic purposes of Phase I was to identify the variables which would predict the observation score given a teacher by an observer-judge. Using the Teacher Practices Observation Record scores as responses, multiple regression models were developed to isolate important variables that would be useful in explaining an observation score given by an observer based on information available about the observer and the filmed teaching situation he observed. Thus, an attempt was made to describe the score given by an observer as a linear function of the variables describing the observer and the filmed teaching episode. A regression line was fitted by method of least squares using the observer's score as a function of these descriptive variables.

Table 11 is a description of the variables considered in this investigation.

Table 11

Definition of Variables

Name of Variable	Statistical Variable	Definition
Response	y_1	TPOR score
Films	x_1	1 if film #2, 0 otherwise (film #1 in base line)
	x_2	1 if film #3, 0 otherwise
	x_3	1 if film #4, 0 otherwise
	x_4	1 if film #5, 0 otherwise
Belief Scores	x_5	Personal Beliefs Inventory
	x_6	Teacher Practices Inventory
	x_7	Personal Opinion Questionnaire
Subject Matter Field of Observer	x_8	1 if Soc. St., 0 otherwise (elementary in the base line)
	x_9	1 if Nat. Sci., 0 otherwise
	x_{10}	1 if Math, 0 otherwise
	x_{11}	1 if Eng. or For. Lang., 0 otherwise
	x_{20}	1 if Generalist, 0 otherwise
Occupational Classification of Observer	x_{12}	1 if Methods Prof., 0 otherwise (clinical sup. in base line)
	x_{13}	1 if Education Prof., 0 otherwise
	x_{14}	1 if Academician, 0 otherwise
Sex	x_{15}	1 if female, 0 if male
Age	x_{16}	chronological age
Institution	x_{17}	1 if Northwestern, 0 otherwise (U. of Wis. in base line)
	x_{18}	1 if Albany, 0 otherwise
	x_{19}	1 if Sacramento, 0 otherwise

Since the most complex model encountered in this investigation is that for the Teacher Practices Observation Record score, this model will be examined in its symbolic form to clarify its meaning and its use. The B's (betas) represent numerical coefficients, the x's the independent variables and y, the response or score. The model is

$$
\begin{aligned}
y = B_0 + B_1 x_1 + B_2 x_2 + B_3 x_3 + B_4 x_4 &\qquad \text{Films}\\
+ B_{17} x_{17} + B_{18} x_{18} + B_{19} x_{19} &\qquad \text{Institution}\\
+B_{12} x_{12} + B_{13} x_{13} + B_{14} x_{14} &\qquad \text{Occupation}\\
+ B_5 x_5 + B_6 x_6 + B_7 x_7 &\qquad \text{Belief Scores}
\end{aligned}
$$

Now x_1, x_2, x_3, and x_4 describe the films; x_{17}, x_{18}, and x_{19} describe the institution; and x_{12}, x_{13}, and x_{14} describe the occupation of the judge. Each of these qualitative variables takes on the value one if the variable describes the judge or the film of interest, and the value zero otherwise. Quantitative variables x_5, x_6, and x_7 represent the judge's belief scores and take on the values of the scores for a given judge.

Suppose that a clinical supervisor from the University of Wisconsin views film one. To predict his rating, we let all the qualitative variables equal zero, which results in the prediction equation

$$y = B_0 + B_5 x_5 + B_6 x_6 + B_7 x_7.$$

Hence, using the judge's belief scores (x_5, x_6, x_7) we would have a predicted value, y, for the judge's rating (within bounds of error).

Now, suppose an academician from Albany views film three. To predict his rating, we let $x_2 = 1$, $x_{14} = 1$, and $x_{18} = 1$, setting all the other qualitative variables equal to zero. The prediction equation now becomes

$$y = (B_0 + B_2 + B_{14} + B_{18}) + B_5 x_5 + B_6 x_6 + B_7 x_7$$

which will yield a predicted value, y, of the judge's rating when the belief scores are substituted for x_5, x_6, and x_7 (within bounds of error).

With this prediction equation,[7] one can predict the rating given any film by a judge, provided one knows his belief scores, his academic position, and the university he represents.

---

[7] William Mendenhall, An Introduction to Linear Models and the Design and Analyses of Experiments (Belmont, California: Wadsworth Publishing Co., Inc., 1966).

In the model search, a detailed model (interaction) was first considered for the response (Teacher Practices Observation Record score) and investigated. It was found that including the interaction terms did not increase the predictive value of the model. See Table 12.

A simpler model (main effect model) was investigated next in an attempt to isolate those variables most useful in predicting an observer's score. A stepwise regression was performed, using the variables films, scores, sex of the observer, age, subject matter field of the observer, occupational classification, and institutional affiliation. At any given step in this regression program, the reduction in the total variation of the ratings accounted for by regression on those variables entered into the regression is computed and tabulated. As each additional variable is entered, the additional reduction in the total variation is computed and tested for statistical significance. Typically, the order in which the variables are entered into the regression becomes an important factor and hence five different orderings considered by the experimenter to be of importance were investigated. In every case, the orderings yielded the same variables as being important. Two orderings will be reported here, namely, the ordering in which the variable "films" is entered first, and the ordering in which "films" is entered last so that one can see the strong overriding influence that "films" has upon a judge's rating.

Findings - TPOR Scores. In the investigation of the TPOR scores the variables found to contribute to the accuracy of prediction were films, institutions, occupational classification, and belief scores. Tables 13 and 14 give one a basis for making this decision. It is well to note the statistical significance of the variable films regardless of the order in which it enters the model. The final regression equation fitted to those variables considered most important in prediction is given in Table 15, together with the Analysis of Variance and Summary Table. The symbol R^2 represents that fractional part of the total variability of the judges' ratings accounted for by regression. It is interesting to note that having fitted only the variable films, R^2 is .4330. Adding the other 3 variables (institutions, occupations, and belief scores) R^2 increases to .4799, showing that the other variables accounted for only 5% more of the variability.

Summary - TPOR Multiple Regression Models. Forty-three percent of the variance in the Teacher Practices Observation Record scores is accounted for by the filmed teaching episodes; the differences between the filmed teacher behavior accounts for this much of the variance in the scores recorded by the observers.

With the addition of the other variables in this model, (belief scores, occupation and institutions of the observers) the total

Table 12

Interaction Model - TPOR

Summary of F-tests to Determine Significance of Variables in the Model

Variables Fitted	$SSR_2 - SSR_1$	$d.f._1$	$\frac{SSR_2 - SSR_1}{d.f._1}$	$d.f._2$	$\frac{SSE}{d.f._2}$	F Value
Main Effects	139,624.236	19	7,348.644	527	281.214	26.132**
Films x Belief Scores	3,644.522	12	303.710	515	280.690	1.082
Films x Sex	1,696.654	4	424.163	511	279.567	1.517
Films x Age	1,098.684	4	274.671	507	279.605	<1.0
F x Subj. Matter & Age x Belief Scores	4,634.074	23	201.481	484	283.318	<1.0
Age x Subj. Matter	1,599.478	4	399.870	480	282.347	1.416
Interactions	12,673.412	47	269.647	480	282.347	< 1.0

** indicates significance at .01 level
* " " " .05 "

Table 13

Summary of F-tests to Determine Significance of Variables in the Model

TPOR - Order 1

Variables Fitted	SSR_2-SSR_1	d.f.$_1$	$\frac{SSR_2 - SSR_1}{d.f._1}$	d.f.$_2$	$\frac{SSE}{d.f._2}$	F Value
Films	124,639.635	4	31,159.909	542	301.078	103.49**
Belief Scores	3,024.550	3	1,008.183	539	297.143	3.39
Subject Matter	1,346.999	5	269.400	534	297.407	<1.0
Institution	8,247.703	3	2,749.234	531	283.555	9.696*
Occupational Classification	1,988.266	2	994.133	529	280.868	3.54
Sex	89.682	1	89.682	528	281.230	<1.0
Age	289.72	1	289.72	527	281.214	1.030

** denotes significance at .01 level
* " " " .05 "

Table 14

Summary of F-tests to Determine Significance of Variables in the Model

TPOR - Order 3

Variables Fitted	$SSR_2 - SSR_1$	$d.f._1$	$\frac{SSR_2 - SSR_1}{d.f._1}$	$d.f._2$	$\frac{SSE}{d.f._2}$	F Value
Belief Scores	2,610.684	3	870.228	543	525.255	1.657
Sex	91.535	1	91.535	542	526.055	<1.0
Age	476.265	1	476.265	541	526.147	<1.0
Subject Matter	1,945.297	5	389.059	536	528.314	<1.0
Occupational Classification	3,285.199	2	1,642.600	534	524.141	3.134*
Institution	9,431.840	3	3,143.950	531	509.340	6.172**
Films	122,259.658	4	30,564.912	527	281.214	108.689**

* denotes significance at .05 level
** " " " .01 "

Table 15

Final Model for TPOR

Equation

$$\begin{aligned} y_1 = {} & 90.8766 + 35.6526\, x_1 + 40.9784\, x_2 + 23.5528\, x_3 + 21.5150\, x_4 \\ & + 2.1853\, x_{17} + 3.0010\, x_{18} + 10.8685\, x_{19} \\ & - 8.1203\, x_{12} - 3.5617\, x_{13} - 1.6493\, x_{14} \\ & - 0.1020\, x_5 + 0.0514\, x_6 - 0.0326\, x_7 \end{aligned}$$

Analysis of Variance

Source	df	ss	MS	F
Regression	13	138,128.961	10,625.305	37.832**
Error	533	149,695.113	280.854	
s = 16.7587				

**denotes significance at .01 level

Summary

Variables	R^2	Inc. in R^2
Films	.4330	.4330
Films, Institution	.4615	.0285
Films, Institution & Occupation	.4705	.0090
Films, Institution, Occupation, & Belief Scores	.4799	.0094

percentage of variance accounted for is forty-eight percent. This indicates the additional variables add only five percent to the total amount of variability that can be identified. Thus, fifth-two percent of the variance from observation scores is unaccounted for and must be explained by chance, error or unmeasured variables.

The additional variables that were included in the models (age, sex, and subject matter field of the observer) seemed to have no statistically significant influence on observer scores, at least in this situation.

From the analysis of the TPOR scores in the first phase of the study it can be tentatively concluded that there are many powerful factors affecting the perceptions of the observers in recording classroom behavior that have not been identified. However, the data indicated that the *Teacher Practices Observation Record* has substantial power for distinguishing differences among the five filmed teachers.

Classroom Behavior Rating Scale

Analysis of Data. The data from the first phase of the study was also analyzed to predict the score of the *Classroom Behavior Rating Scale* given the filmed teacher behavior by the observer-judge. The same statistical procedures (multiple regression models) were used to identify predictor variables of the CBRS scores as were used for the *Teacher Practices Observation Record* scores (see page 19). The variables considered for this investigation were the same as those considered for the TPOR and are reported in Table 11.

The *Classroom Behavior Rating Scale* consists of seventeen items which describe teacher and pupil behavioral characteristics. Thirteen of the items describe teacher characteristics, *Classroom Behavior Rating Scale-Teacher* (CBRST) and the remaining four items describe pupil characteristics, *Classroom Behavior Rating Scale-Pupil* (CBRSP). For the purposes of this analysis the CBRST and CBRSP scores were first considered independently and then combined. Thus, three responses--CBRSP, CBRST, and CBRS--were used to isolate predictor variables.

In the model search an interaction model was first considered for the three responses. It was found in each case that including the interaction terms did not increase the predictive value of the models. See Table 16.

Next, main effect models were investigated to identify those variables most useful in predicting an observer's rating. Again stepwise regressions were performed using the variables of films, belief scores, sex, age, and subject matter field of the observer, occupational classification and institutional classification.

Findings - CBRS Rating Scores. In investigating the CBRSP ratings, the variables films, belief scores, and institutions were found to be most useful in the prediction of the observer's rating. Tables 17 and 18 display the results of the two orderings, again

Table 16

Interaction Model - CBRSP

Summary of F-tests to Determine Significance of Variables in the Model

Variables Fitted	$SSR_2 - SSR_1$	$d.f._1$	$\frac{SSR_2 - SSR_1}{d.f._1}$	$d.f._2$	$\frac{SSE}{d.f._2}$	F Value
Main Effects	9,386.766	20	469.338	526	88.548	5.300**
Films x Belief Scores	842.421	12	70.200	514	88.977	<1.0
Films x Sex		4	75.557	510	89.457	<1.0
Films x Age	220.031	3	73.344	506	89.351	<1.0
Films x Subj. Matter		20	93.538	486	89.179	1.050
Sex x Belief Scores		3	70.343	483	89.296	<1.0
Institution x Occupation	1,037.485	9	115.276	474	88.803	1.298
Interactions	4,483.000	51	87.92	474	88.803	<1.0

** indicates significance at .01 level
* " " " .05 "

Table 17

Summary of F-tests to Determine Significance of Variables in the Model
CBRSP - Order 1

Variables Fitted	$SSR_2 - SSR_1$	d.f.$_1$	$\frac{SSR_2 - SSR_1}{d.f._2}$	d.f.$_2$	$\frac{SSE}{d.f._2}$	F Value
Films	5,394.882	4	1,348.721	542	91.856	14.68**
Belief Scores	2,068.73	3	689.577	539	88.529	7.789*
Sex	93.635	1	93.635	538	88.519	1.058
Age	14.392	1	14.392	537	88.657	<1.0
Subject Matter	383.699	5	76.7398	532	88.972	<1.0
Occupation	387.890	2	193.945	530	88.576	2.190
Institution	206.693	3	86.897	527	88.586	<1.0

** denotes significance at .01 level
* " " " almost .05 level

Table 18

Summary of F-tests to Determine Significance of Variables in the Model
CBRSP - Order 3

Variables Fitted	$SSR_2 - SSR_1$	$d.f._1$	$\frac{SSR_2 - SSR_1}{d.f._1}$	$d.f._2$	$\frac{SSE}{d.f._2}$	F Value
Scores	1,982.576	3	660.859	543	97.971	6.745**
Institution	485.352	3	161.784	540	97.616	1.657
Occupation	448.377	3	149.459	537	97.326	1.536
Subject Matter	208.953	4	52.238	533	97.665	<1.0
Sex	25.996	1	25.996	532	97.799	<1.0
Age	33.824	1	33.824	531	97.920	<1.0
Films	5,310.816	4	1,327.704	527	88.586	14.988**

** denotes significance at .01 level

indicating that films are perhaps the most important variable to be considered. The final fitted regression equation using only films, belief scores, and positions is given in Table 19, together with an A.O.V. and Summary Table. Considering only the variable films, $R^2 =$ .0978, whereas including the variables belief scores and positions increases R^2 to .1439 to further point out the dominance of the films as a predictor variable.

Since R^2 remains so low, one infers that the true variance of the ratings is quite high and decreased very little by taking account of the variables we have measured.

The variables films, belief scores, and institutions were found to be the important variables in the prediction of the CBRST rating (see Tables 20 and 21). Using these variables, the final prediction equation was fitted and is given in Table 22. Using films as the only variable, one finds that R^2, the fractional part of the total variation in the ratings is .1357. By adding the variables positions and scores, R^2 increases to .1791. Hence, of the three variables considered to be of importance, films again is the most useful. The low value of R^2 leads one to believe that the true variability of the ratings is quite high and decreased very little by taking account of the variables measured.

The CBRS rating is a combination of the CBRST and CBRSP ratings; therefore, one would expect to arrive at the same conclusions for this rating as for the CBRST and CBRSP. This is indeed the case. Films, occupation, and belief scores are the best predictor variables for the CBRS rating (see Tables 23 and 24). When the prediction equation is fitted (see Table 25) to these three variables, R^2 is found to be .1691, an increase of .0510 from the value $R^2 = .1181$ obtained by fitting films only.

It is noted that the high variability inherent in the CBRST and CBRSP is in evidence in the combined CBRS rating in the form of a low value of R^2.

Summary - CBRS Multiple Regression Model. Only twelve percent of the variance in the Classroom Behavior Rating Scale scores were accounted for by the filmed teaching episodes. Adding the other statistically significant variables, belief scores and occupational classification of the observers, only increased the variance that can be identified to seventeen percent of the total variance. In contrast to the Teacher Practices Observation Record for which forty-eight percent of the variance in scores can be identified, the CBRS seems much less a reflection of the actual teaching behavior seen in the films. Although a large part of the variance in the observation scores is due to unisolated variables and error, the filmed teaching behavior has a much greater effect on observation scores than on the rating scores. This would indicate that the observer-judges, in this situation at least, rate teachers on many criteria other than their professed beliefs about education and the actual classroom behavior of the teacher.

Table 19

Final Model CBRSP

Equation

$$y_2 = 22.7161 + 7.5939\ x_1 - 8.1812\ x_2 - 4.3280\ x_3 - 2.7585\ x_4$$
$$+ 0.0659\ x_5 - 0.0203\ x_6 + 0.0292\ x_7$$
$$+ 3.78683\ x_{12} + 1.4535\ x_{13} - 0.1421\ x_{14}$$

Analysis of Variance

Source	df	ss	MS	F
Regression	10	7,940.885	794.089	9.010*
Error	536	47,239.698	88.134	
s = 9.3879				

*denotes significance at .05 level

Summary

Variables	R^2	Inc. in R^2
Films	.0978	.0978
Films and Occupation	.1178	.0200
Films, Occupation & Belief Scores	.1439	.0261

Table 20

Summary of F-tests to Determine
Significance of Variables in the Model
CBRST - Order 1

Variables Fitted	$SSR_2 - SSR_1$	$d.f._1$	$\frac{SSR_2 - SSR_1}{d.f._1}$	$d.f._2$	$\frac{SSE}{d.f._2}$	F Value
Films	1,035.863	4	258.966	542	12.177	21.267**
Belief Scores	285.084	3	95.028	539	11.696	8.125[a]
Sex	.671	1	.671	538	11.736	<1.0
Age	1.670	1	1.670	537	11.755	<1.0
Subject Matter of Observer	69.645	5	13.929	532	11.738	1.187
Occupation	105.090	2	52.545	530	11.584	4.536
Institution	13.754	3	4.585	527	11.624	<1.0

** denotes significance at .01 level
a " " " .05 "

Table 21

Summary of F-tests to Determine Significance of Variables in the Model
CBRST - Order 3

Variables Fitted	$SSR_2 - SSR_1$	$d.f._1$	$\frac{SSR_2 - SSR_1}{d.f._1}$	$d.f._2$	$\frac{SSE}{d.f._2}$	F Value
Scores	283.234	3	94.411	543	13.541	6.972^{**}
Sex	1.584	1	1.584	542	13.563	<1.0
Age	2.940	1	2.940	541	13.582	<1.0
Subject Matter of Observer	58.749	5	11.750	536	13.599	<1.0
Occupation	90.977	2	45.489	534	13.480	3.375^{*}
Institution	9.784	3	3.261	531	13.538	<1.0
Films	1,062.839	4	265.710	527	11.624	22.859^{**}

** denotes significance at .01 level
* " " " .05 "

Table 22

Final Model CBRST

Equation

$$y_3 = 8.1052 - 2.9417\ x_1 - 3.9550\ x_2 - 1.9296\ x_3 - 2.5529\ x_4$$
$$+ 0.0282\ x_5 - 0.0062\ x_6 + 0.0027\ x_7$$
$$+ 1.3782\ x_{12} + 0.4116\ x_{13} + 0.1595\ x_{14}$$

Analysis of Variance

Source	df	ss	MS	F
Regression	10	1,367.681	136.768	11.695**
Error	536	6,268.098	11.694	
s = 3.4196				

**denotes significance at .01 level

Summary

Variables	R^2	Inc. in R^2
Films	.1357	.1357
Films & Occupation	.1500	.0143
Films, Occupation & Belief Scores	.1791	.0291

Table 23

Summary of F-tests to Determine Significance of Variables in the Model
CBRS - Order 1

Variables Fitted	$SSR_2 - SSR_1$	$d.f._1$	$\frac{SSR_2 - SSR_1}{d.f._1}$	$d.f._2$	$\frac{SSE}{d.f._2}$	F Value
Films	10,816.573	4	2,704.143	542	148.972	18.152**
Belief Scores	3,860.215	3	1,286.738	539	142.640	9.021*
Sex	78.460	1	78.460	538	142.759	<1.0
Age	25.867	1	25.867	537	142.977	<1.0
Subject Matter of Observer	416.661	5	83.332	532	143.537	<1.0
Occupation	763.848	2	381.924	530	142.638	2.678
Institution	275.046	3	91.682	527	142.659	<1.0

** denotes significance at the .01 level
* " " " " .05 "

Table 24

Summary of F-tests to Determine Significance of Variables in the Model
CBRS - Order 3

Variables Fitted	$SSR_2 - SSR_1$	$d.f._1$	$\frac{SSR_2 - SSR_1}{d.f._1}$	$d.f._2$	$\frac{SSE}{d.f._2}$	F Value
Belief Scores	3,712.256	3	1,237.419	543	161.782	7.649**
Sex	57.619	1	57.619	542	161.974	<1.0
Age	34.823	1	34.823	541	162.209	<1.0
Subject Matter	345.976	5	69.195	536	163.076	<1.0
Occupation	774.421	2	387.210	534	162.237	2.387
Institution	409.719	3	136.573	531	162.382	<1.0
Films	10,901.853	4	2,725.463	527	142.928	19.069**

** denotes significance at the .01 level

Table 25

Final Model CBRS

Equation

$$\begin{aligned} y_4 = {} & 30.8213 - 10.5357\ x_1 - 12.1362\ x_2 - 6.2576\ x_3 - 5.3114\ x_4 \\ & + 0.0941\ x_5 - 0.0265\ x_6 + 0.0319\ x_7 \\ & + 5.1650\ x_{12} + 1.8651\ x_{13} + 0.0174\ x_{14} \end{aligned}$$

Analysis of Variance

Source	df	SS	MS	F
Regression	10	15,481.511	1,548.151	10.907**
Error	536	76,078.139	141.937	
s = 11.9137				

**denotes signficance at .01 level

Summary

Variables	R^2	Inc. in R^2
Films	.1181	.1181
Films & Occupation	.1388	.0207
Films, Occupation & Belief Scores	.1691	.0303

CHAPTER IV

PHASE II - OBSERVER-JUDGE RATINGS OF STUDENT TEACHERS

Purposes

The difficulty in making meaningful judgments of teacher competence centers squarely on obtaining hard evidence about what teachers and their students do in the classroom. Therefore, this study has two major purposes:

1. To identify variables and combinations of variables which contribute significantly to variance in observations of the classroom behavior of student teachers.

2. To identify variables (including observation scores) which contribute significantly to variance in the evaluation of classroom behavior.

Procedures

A total of 569 observer-judges made 2,859 observations and 953 evaluations of 407 student teachers from six teacher education institutions in California, Florida, Illinois, New York, and Wisconsin. Observations of the classroom behavior of the students were made with the *Teacher Practices Observation Record* (TPOR), a 62-item sign system which measures congruity of teaching methods with John Dewey's philosophy of experimentalism. Judgments with respect to the quality of observed classroom behavior were made on the *Teacher Evaluation Scale* (TES), which is a simple form for rating teachers along a six-point competent-incompetent continuum on six general teacher characteristics. In addition, scores on the *Personal Beliefs Inventory* (PBI) and the *Teacher Practices Inventory* (TPI), which measure congruity of beliefs with Dewey's experimentalism were obtained for each student teacher and observer-judge in an effort to assess the influence of their personal and educational philosophy on the observations and evaluations.

The student teachers in this study were drawn from those engaged in their final pre-service clinical experience under the auspices of the six cooperating colleges and universities during the winter and spring terms of 1965. The observer-judges consisted of student teacher supervisors, education professors, and academic professors from all six colleges, plus cooperating teachers and principals from the public schools participating in the regular teacher education programs of these six institutions.

Analysis of Data - Using Total Scores of Observations and Evaluations

The data were submitted to multiple regression analysis to discover what measured variables contributed to the scores given by the observer-judges on the Teacher Practices Observation Record (TPOR) and the Teacher Evaluation Scale (TES). Each of these two scores were treated separately, and in turn, as the predicted response in a series of three increasingly complex regression models, first using 12 independent variables, then 20, and finally a total of 69, including two-way interactions of these variables.

Several models were proposed in order to determine which variables contribute most to each response. The limitations due to the size of the computer memory banks and the tremendous number of variables involved in any evaluation were realized; independent variables could be introduced only as linear factors, and interactions between the variables were explored only on a limited basis. The primary objective was to identify variables making significant contributions to the observation and evaluation scores.

"Significance" was determined in two ways:

1) The significance of the regression itself is tested by a simple analysis of variance. Consider the general linear model:

   $$Y = \beta_0 + \beta_1 X_1 + \ldots. + \beta_k X_k + \epsilon$$

   If it is desired to know whether the k independent variables account for a significant amount of response variance, we perform the test:

   H_0: the model makes no contribution, i.e.,

   $\beta_1 = \beta_2 = \ldots. = \beta_k = 0$

   vs. H_a: the model makes a significant contribution.

   It can be shown that under the null hypothesis, H_0:

   $$F = \frac{SS\ (\text{Regression})/k}{SS\ (\text{Error})/(n - (k + 1))}$$

   has an F distribution. We thus reject H_0, i.e., conclude that the model does make a significant contribution in accounting for the variance of the response, if the calculated F value exceeds the table F (α, ν_1, ν_2), where α is the level of significance, ν_1 is the degrees of freedom for the regression (k in above example), and ν_2 is the degrees of freedom associated with error ($n - (k + 1)$ above).

2) Once significance of regression has been determined, those variables which contribute most or contribute significantly must be determined. Suppose we wish to know whether

variable X_i is significant—we then perform the test:

$$H_0: \ \beta_i = 0$$

$$H_a: \ \beta_i \neq 0$$

using $t = \dfrac{\hat{\beta}_i}{S_{\hat{\beta}_i}}$, where

$\hat{\beta}_i$ is the estimator of the coefficient β_i, and $S_{\hat{\beta}_i}$ is its standard deviation, as the test statistic. Here, t is distributed as Student's t, and if $|t| > t\ (\alpha, \nu_2)$ the conclusion is that variable i makes a significant contribution.

Certain disadvantages are inherent using the above method of analysis. There exists an increasingly large probability of making an incorrect decision as the number of variable coefficients tested in this manner increases, since the probability of rejecting H_0 when it is true is α for each test. But this is irrelevant to our objective—the aim of the study is to screen out those independent variables whose contribution seems to be insignificant. That we might include as significant some which, in fact, make little or no contribution is rather unimportant since each of these will be explored in greater depth, and, if not screened by this regression, will likely be caught in future analysis.

Thus, variable coefficients whose calculated t-values are even close to being significant are noted for further study, and if a group of "dummy" variables is used to describe one particular factor, e.g., the five variables which describe the six institutions at which observations originated, and include several which have significant or nearly significant t-values, the factor is "screened," and its contribution considered significant.

Table 26 is a description of the variables considered in the investigation.

I. Response: TPOR Total Score

Model 1: The general model was fitted with 10 independent variables:

PBI TPI } observer	2 variables
PBI TPI } student-teacher	2 variables
6 two-way interactions of TPI and PBI of student-teacher and observer	6 variables

Table 26

Definition of Variables

Factor	No. of Variables	Description
Belief Scores of Observer	2	Personal Beliefs Inventory of Observer Teacher Practices Inventory of Observer
Belief Scores of Student Teacher	2	Personal Beliefs Inventory of Student Teacher Teacher Practices Inventory of Student Teacher
Occupational Classification of Observer	4	Cooperating Teacher Principal Clinical Professor Methods Professor or Educ. Prof. Academician
Institutional Affiliation	5	Sacramento Albany Oneonta Wisconsin Northwestern Florida
Subject Matter of Observer	1	Secondary, Elementary
Age of Observer	1	---
Sex of Observer	1	Female, Male
Subject Matter of Student Teacher	1	Secondary, Elementary
Age of Student Teacher	1	---
Sex of Student Teacher	1	Female, Male

2

The F-value was 9.78, which is significant at the .01 level. The regression model contributed a significant amount of information about the TPOR scores.

Model 2: The general model was fitted with 19 independent variables. They included the following:

Position or occupation of observer		4 vars.
Institution at which observation took place		5 vars.
Subject (Elem. or Sec.) Age Sex	of observer	3 vars.
Subject Age Sex	of student teacher	3 vars.
PBI TPI	observer	2 vars.
PBI TPI	student teacher	2 vars.

All variables were included as linear terms, and no interactions were included.

The calculated F-value was 9.17, again significant at the .01 level. Thus, the model accounts for a significant amount of response variance.

Model 3: In addition to the above 19 independent variables, Model 3 included all 2-way interactions which were considered pertinent. Limitations of the computer made it impossible to consider all 2-way interactions of these 19 variables, but the following 50 were included:

Position x Sex - Observer	4 vars.
Position x Sex - Student Teacher	4 vars.
Position x PBI - Observer	4 vars.
Position x TPI - Observer	4 vars.
Institution x PBI - Student Teacher	5 vars.

Institution x TPI - Student Teacher	5 vars.
Subject - Student Teacher x Subject - Observer	1 var.
Subject - Judge x PBI and TPI - Observer	2 vars.
Age - Judge x Age - Student Teacher	1 var.
Age - Judge x PBI and TPI - Observer	2 vars.
Sex - Judge x Sex and Age - Student Teacher	2 vars.
Sex - Judge x PBI and TPI - Observer	2 vars.
Age - Student Teacher x Sex and Subject - Student Teacher	2 vars.
Age - Student Teacher x PBI and TPI - Student Teacher	2 vars.
Sex - Student Teacher x PBI and TPI - Student Teacher	2 vars.
Subject - Student Teacher x PBI and TPI - Student Teacher	2 vars.
All 2-way interactions of PBI and TPI of Student Teacher and Observer	6 vars.

An F-value of 4.49 resulted, significant at the .01 level. In addition, to test the significance of the interaction in terms as a group, the following test was conducted.

H_0: the interaction terms are insignificant, i.e.,

$\beta_{20} = \beta_{21} = ... = \beta_{69} = 0$

where $\beta_{20} - \beta_{69}$ represent the 50 coefficients of the interaction terms.

H_a: the interactions make a significant contribution.

Test Statistic:

$$F = \frac{(SSE_2 - SSE)/k_2 - k_1}{SSE_1/k_1}$$

where SSE_2 and k_2 are the sums of squares due to error and degrees of freedom, respectively, associated with error in the reduced model (no interactions), and SSE_1 and k_1 are similar notations for the complete model (interactions included).

Then $$F = \frac{(3895.16 - 3431.83)/(968 - 918)}{3431.83/918}$$

$$= 2.48$$

Thus the interactions considered as a group make a significant contribution in accounting for response variance.

II. Response: Evaluations of Student Teachers

Model 1: The general model was fitted with 15 independent variables. These included the following:

TPOR total score	1 var.
PBI } Judge and TPI } Student Teacher	4 vars.
All possible 2-way interactions of above	10 vars.

The model was highly significant in its contribution to response variation, with an F-value of 25.63.

Model 2: The general model was fitted with 20 independent variables, all introduced in linear form with no interactions included:

Position of Judge	4 vars.
Institution	5 vars.
Subject, Age, and Sex of Judge	3 vars.
Subject, Age, and Sex of Student Teacher	3 vars.
TPOR score	1 var.
PBI, TPI of Judge and Student Teacher	4 vars.

The F-value was calculated to be 19.33, implying the regression model to be highly significant.

Model 3: The general model was fitted with the 20 independent variables above plus 69 two-way interactions of these variables. The interactions included were:

Position x Sex - Student Teacher	4 vars.

Position x Sex - Judge	4 vars.
Position x TPOR	4 vars.
Position x Beliefs - Judge	8 vars.
Institution x TPOR	5 vars.
Institution x Beliefs - Student Teacher	10 vars.
Subject - Judge x Subject - Student	1 var.
Subject - Judge x TPOR	1 var.
Subject - Judge x Beliefs - Judge	2 vars.
Age - Judge x TPOR	1 var.
Age - Judge x Beliefs - Judge	2 vars.
Sex - Judge x TPOR	1 var.
Sex - Judge x Beliefs - Judge	2 vars.
Sex - Judge x Sex - Student Teacher	1 var.
Sex - Judge x Age - Student Teacher	1 var.
Age - Student Teacher x Sex - Student Teacher	1 var.
Age - Student Teacher x Subject - Student Teacher	1 var.
Age - Student Teacher x TPOR	1 var.
Age - Student Teacher x Beliefs - Student Teacher	2 vars.

Sex - Student Teacher x TPOR	1 var.
Sex - Student Teacher x Beliefs - Student Teacher	2 vars.
Subject - Student Teacher x TPOR	1 var.
Subject - Student Teacher x Beliefs - Student Teacher	2 vars.
Age - Judge x Age - Student Teacher	1 var.
All 2-way interactions of TPOR, and PBI, TPI of Judge and Student Teacher	10 vars.

An F-value of 6.11 resulted from the use of the above model, showing the regression's contribution significant at the .01 level. Also, the test of interaction significance was run to determine whether the interaction terms, considered as a whole, make a contribution in accounting for evaluation variance. An F-value was calculated as follows:

$$F = \frac{(SSE_2 - SSE_1/(k_2 - k_1)}{SSE_1/k_1}$$

$$= \frac{(18979.86 - 16544.45)/(967 - 898)}{16544.45/898}$$

$$= 1.91$$

Thus, the interactions' contribution is significant at the .01 confidence level.

Findings of Total Score Analysis

I. Variables which contribute significantly to the variance in observations of the classroom behavior of student teachers as measured by scores on the TPOR are shown in Table 27.

II. Variables which contribute significantly to the variance in evaluations of teaching competence as measured by TES scores are shown in Table 28.

Discussion of Total Score Analysis

I. What variables and combinations of variables contribute significantly to the variance in observations of the classroom behavior of student teachers as measured by scores on the TPOR?

Table 27

Variables Which Contribute to Variance in Observation of Classroom Behavior

Response: TPOR Total Scores

	Variable	*t-value
Model 1:	Observer PBI-TPI interaction	-4.90
	Student Teacher TPI-Observer PBI interaction	1.97
Model 2:	Student Teacher TPI	4.29
	Student Teacher PBI	-2.35
	Age of Observer	4.37
	Occupation of Observer (4)	4.37 to -5.42
Model 3:	Observer Age x St. Teacher Age	-1.76
	Observer Age x Observer TPI	2.19
	Observer Sex x Observer TPI	2.45
	St. Teacher Sex x St. Teacher PBI	-2.23
	St. Teacher PBI x St. Teacher TPI	1.53
	Observer PBI x Observer TPI	-2.39
	Observer Sex x Observer Occupation (4)	2.51 to -3.63
	Observer PBI x Observer Occupation (4)	1.62 to -2.28
	Observer TPI x Observer Occupation (4)	2.70 to -2.24
	Institution x St. Teacher TPI (5)	2.34 to -1.81

*t-value = 1.96, significant at .05 level
t-value = 1.645, significant at .10 level

Table 28

Variables Which Contribute to Variance In Evaluations of Teaching Competence

Response: TES Total Scores

	Variable	*t-value
Model 1:	TPOR x Judge TPI	2.97
	Judge PBI x Judge TPI	-3.88
	Judge PBI x St. Teacher TPI	1.98
	Judge TPI x St. Teacher PBI	1.75
Model 2:	TPOR	13.44
	St. Teacher Sex	-3.67
	Judge TPI	-2.63
	Judge PBI	-1.94
	Student Teacher Age	2.23
	Judge Occupation (4)	2.21 to -2.43
	Institution (5)	3.41 to -2.85
Model 3:	TPOR x Judge TPI	2.24
	TPOR x St. Teacher PBI	-1.99
	Judge PBI x St. Teacher PBI	-1.85
	Judge TPI x St. Teacher PBI	1.62
	Judge Subject x Judge TPI	-1.80
	Judge Sex x St. Teacher Sex	3.26
	St. Teacher Age x St. Teacher Subject	2.71
	St. Teacher Sex x St. Teacher TPI	-1.81
	Judge Sex x Judge Occupation (4)	2.09 to -2.00

*t-value = 1.96, significant at .05 level
t-value = 1.645, significant at .10 level

A. How do the beliefs of the student teacher influence the TPOR score received? When the PBI and TPI of the student teacher are introduced linearly in Model 2, their relations to the TPOR score received seem to be conflicting in nature. While the TPI is in direct proportion to the TPOR score, the PBI seems to be inversely proportional to the TPOR. This would seem to deliver a damaging blow to the value of the PBI as a predictor of observed behavior--except that we have had repeated experience with the "hidden power" of the PBI total score before and know it for what it is--a real sleeper.

Model 1 indicates that high PBI-TPI scores of the observer-judge make the positive effect of a high student teacher TPI more pronounced, as does a high student teacher PBI. If a student teacher who has a high TPI also has a high PBI the probability of his receiving a high TPOR is sharply increased. Thus, the PBI acts something like an "additive" in the gas tank of the high TPI student teacher. It provides the "extra kick" that makes the difference. Furthermore, as these three scores (student teacher PBI, and the PBI-TPI of the observer-judges) decrease the student teacher's TPI relation to the TPOR decreases, and may even become inverse. Clearly, the evidence indicates that the complex relationships within the chemistry of the beliefs of both the student teacher and his observer exercise an influence upon the observation (TPOR) score which cannot be ignored.

B. How do the beliefs of the observer-judge influence the TPOR scores given? The PBI and TPI of the observer-judge seem to enjoy a more straightforward relationship with the TPOR scores. The PBI of the observer-judge is inversely proportional to the TPOR, with this effect seemingly becoming more pronounced as the TPI of the observer-judge increases. As the TPI of the student teacher being observed increases the relationship between the observer-judge's PBI and the TPOR becomes less inverse. In other words, it takes a student teacher with a high TPI to break down the propensity of a high PBI observer-judge for giving low TPOR scores. The TPI of the observer-judge also appears to be inversely proportional to the TPOR given, but this effect is very slight unless the PBI of the observer-judge and the PBI of the student teacher are high.

C. How do the beliefs of the student teacher and the observer-judge interact with other descriptive factors to influence TPOR scores? When the beliefs scores of the students and observers and their interactions with various other variables were examined in Model 3, the TPI of the student teacher did not vary significantly in the nature of its previously stated relationships. A slightly significant interaction was found between his TPI and the institution he attended, but this seems more relevant to a study aimed at the comparative evaluation of the cooperating institutions than to the central purposes of our investigation.

On the other hand, the PBI of the student teacher, which appeared in Model 1 to be negatively related to the TPOR, appears in Model 3 to act in conjunction with several other factors to relate positively with the TPOR. The student teacher's PBI tends to become positively correlated to the TPOR as his TPI score increases. This effect is even more pronounced if the teacher is female, and/or if the observer-judge has a low PBI score. Remember, low PBI observers tend generally to give high TPOR scores. They have difficulty differentiating experimental from non-experimental behavior in the classroom because they neither understand nor appreciate the underlying theoretical dimensions of the TPOR. When they "smell" a student who differs diametrically with them at the level of basic beliefs--a high PBI student teacher--they tend to "punish" him with a high TPOR score. Likewise, females tend generally to be "lukewarm" toward experimentalism, but when one of them decides to throw a high PBI in with a high TPI she becomes "red hot" and goes all the way in her enthusiasm for experimentalism and receives a very high TPOR.

The TPI of the observer-judge generally has an inverse relation with the TPOR. However, the observer-judge with a high TPI tends to give even lower TPOR scores if she happens to be a relatively young female observer-judge with a high PBI score.

A significant relation also seems to exist between the observer belief scores and his position (professional occupation). Detailed analysis of this factor has already been reported elsewhere,[1] showing that college professors (both educationists and academicians) tend toward higher PBI scores than do students, cooperating teachers, principals, and supervisors of student teachers.

Personal characteristics of the observer-judges and student teachers also seem to make certain significant contributions in accounting for TPOR variance. Age of the observer seems to be in direct proportion to the TPOR. As the observer-judge's age increases so does the TPOR score given. This effect is much less pronounced if the student teacher is also older, but more profound if the "old" observer-judge's TPI is high. Older observer-judges--even those with high TPI scores--tend to have low PBI scores. This acts to inflate the TPOR scores given by them, as persons (of any age) who are caught in a serious discrepancy between a high TPI and a low PBI seem to have trouble clearly differentiating experimental from non-experimental behavior on the TPOR.

The most significant contributor among the measured personal characteristics was the sex of the student teacher. Generally, the male teacher seems to receive a higher TPOR than does a female, but this effect is much less pronounced, and possibly even slightly reversed, if the TPI and PBI of the female teacher are high.

II. What variables and combinations of variables contribute significantly to the variance in judgments made with respect to the quality of teaching which was observed?

---

[1] Bob Burton Brown and Tom Rusk Vickery, "The Belief Gap in Teacher Education," *The Journal of Teacher Education*, 18:417-421, Winter, 1967.

A. How does the observation (TPOR) score relate to judgments of teacher competence (TES score)? By far the most significant predictor of the evaluation given the teacher is the TPOR score. This relationship is a very pronounced positive correlation, with the evaluation having an even greater tendency to be favorable if the high TPOR score is accompanied by a high observer TPI, and low student teacher belief scores on both the PBI and TPI. These findings cut in opposite directions: (1) The more the teacher's behavior is observed to be in agreement with Dewey (high TPOR) the more likely it is to be evaluated favorably (high TES), and vice versa, and (2) the lower the student teacher's agreement with Dewey's beliefs the more likely he is to get a favorable rating. However, there is a "catcher" to this statement to which the reader must be made aware. The TPOR scores tended to be low; i.e., on the whole the student teachers did not often employ classroom practices consistent with Dewey's experimentalism. Therefore, relatively non-experimental teaching performances received favorable evaluations, although the more experimental these became the better their rating.

B. How do the beliefs of the observer-judges influence their judgments of teacher competence? The PBI and TPI of the observer-judge seem to have a relationship with evaluation not dissimilar to that with the TPOR. Both are inversely proportional. In short, the higher the observer-judge's expectations with respect to experimentalism (high PBI-TPI) the lower his estimation of the teaching observed (low TES). Likewise, the lower the observer-judge's PBI-TPI, the higher his evaluations. However, the PBI's inverse effect becomes less pronounced if the TPI of both the observer-judge and the student teacher are relatively low.

The TPI of the observer-judge seems to be even more negative than the PBI in its correlation to the evaluation. But the negative effect becomes more emphatic if the TPI is accompanied by a high observer-judge PBI. The effect is somewhat modified if the TPOR score given is relatively high, and/or if the belief scores of the student teacher are relatively high.

One might conclude that the observer-judges expected to observe highly experimental behavior exhibited in the classroom by the student teachers, and, failing to see it, gave them poor ratings. The fact is, almost the opposite happened. Most of the observer-judges turned out to be ambivalent in their beliefs regarding Dewey's experimentalism, observed classroom behavior which was in disagreement about twice as often as it was in agreement with the educational practices advocated by Dewey, and evaluated that teaching on the whole as being "very good" to "excellent."

C. How do the beliefs of the student teachers influence the ratings given their observed teaching behavior? The belief scores of the student teachers seem less significant in their contribution to the evaluation than do those of the observer-judge. Of the two beliefs measures, the TPI of the student teacher is the more significant in predicting the evaluation, with a generally positive correlation between the TPI and the evaluation. This is more pronounced if the TPI of the observer-judge is high, and is modified somewhat if the observer-judge's PBI is high.

The student teacher's PBI seems to enjoy a noticeable positive correlation with the evaluation only if the TPI of the observer-judge is relatively high. Otherwise, the effect appears nil, or even slightly negative. When pertinent interactions are introduced in Model 3, evidence is yielded to indicate that a relatively high TPOR score or a high observer-judge PBI score will tend to modify this effect of the student teacher PBI.

D. How do other descriptive factors influence ratings of teacher competence? Model 3 interactions show that the positive correlation between evaluation and student teacher TPI is more pronounced for a female teacher than for a male teacher.

Possibly, the most interesting of the significant interactions is the one found between sex of the observer-judge and sex of the student teacher. The indication is that the rating given tends to be higher if the observer-judge and the student teacher are of the same sex. Also, an older student teacher seems to receive higher evaluations than a younger one, with this effect becoming even more pronounced for an elementary teacher than for a secondary teacher.

Summary of Total Score Analysis

In summary, the fact that this analysis has chipped only small pieces from a large block of data should be emphasized. That so many important, simple relationships were discovered is highly promising. Areas worthy of further study have been pinpointed, opening the way for the search for many heretofore obscure and subtle relationships involved in the problems of observing and judging classroom behavior. For example, a possible explanation for the often contradictory belief score relationships may be that the various factors within the instruments vary in the contribution to the TPOR and TES ratings given, thus presenting an obscure picture when treated as a single score, as we have done here. This determination of factor significance is reported next.

Re-examination of the Data Using Factor Scores

The Personal Beliefs Inventory (PBI), the Teacher Practices Inventory (TPI), and the Teacher Practices Observation Record (TPOR) are comprised of two different types of factors: (1) "theoretical" factors, and (2) "empirical" factors. The theoretical factors are those which were extracted from the theoretical framework of John Dewey's experimentalism and built into the instruments from the beginning. The empirical factors are those which were identified as the result of submitting the empirical data collected in Phase I to factor analysis. Theoretical and empirical factors are listed and described in Appendix C.

Analysis of Data Using Factor Scores

The theoretical and empirical factors were included in several multiple linear regressions. The purpose of these regressions was to further explore the results of previous regression models by breaking the total scores of the Personal Beliefs Inventory (PBI), Teacher Practices Inventory (TPI), and Teacher Practices Observation Record (TPOR) into factors. In this way it was hoped that some of the conflicting conclusions drawn from the models using only total scores (see the regression analysis in the previous section of this chapter) could be explained.

Four models were proposed. The first two treated the TPOR as the response, one employing the empirical factors of the PBI and TPI as independent variables, the second employing the theoretical factors of those instruments as independent variables. In the last two models the sum of the six Teacher Evaluation Scale (TES) scores was treated as the response; in the first the empirical factors of the PBI, TPI, and TPOR were the independent variables, and in the second the theoretical factors of the same instruments were proposed as independent variables.

In all models the variables were introduced only in a linear manner since the main objective was to determine basic observation-factor relationships; any accurate predictive equation would be incidental. It was found, however, that each of the models accounted for a significant percentage of the variance of the response. Statistical significance of a particular factor was determined by a simple t-test upon its coefficient in the regression equation.

Table 29 lists the multiple R^2 and the F-value resulting from the test of the model's significance.

Table 29

F-Values of Theoretical and Empirical Factors

Response	Factors	R^2	F
TPOR	Theoretical	.1248	3.98**
	Empirical	.1241	5.48**
TES	Theoretical	.3544	12.20**
	Empirical	.3320	14.29**

**Denotes significance at the .01 level.

Findings of the Factor Scores

The tables which follow list the factors which were significant at the .95 level of significance (t = 1.96) in descending order of their t-values. Those which were not significant at the .95 level, but which were at the .90 level (t = 1.645) are starred (*). Included with the name of the factor is the name of the instrument to which the factor belongs, plus an indication whether the score belonged to the student teachers (ST) or the observer-judges (OJ). A minus sign before the t-value indicates that the factor was inversely proportional to the regression response; without a minus sign, the factor score was in direct proportion to the response.

Table 30

Teacher Practices Observation Record (TPOR) Regression

Theoretical Factors

Instrument	Factor		t-value
PBI	T-6	Knowing and Doing (OJ)	-3.77
PBI	T-3	Science and Morals (OJ)	-3.46
TPI	T-11	Mechanical Following of an Established Method (ST)	3.19
PBI	T-2	Change and Certainty (OJ)	2.71
TPI	T-9	Reliance Upon Extrinsic Motivation (OJ)	2.31
TPI	T-2	Development of Challenging Problem (ST)	2.20
TPI	T-1	Situation of Experience (ST)	2.09
PBI	T-4	Emotions and Intellect (ST)	-1.93*
TPI	T-3	Generation of Ideas (ST)	-1.76*

Table 31

Teacher Practices Observation Record (TPOR) Regression

Empirical Factors

Instrument	Factor		t-value
PBI	E-6	Nature of Learning (OJ)	-5.92
TPI	E-1	Evils in Education (OJ)	4.06
PBI	E-1	Science and Morality (OJ)	-3.50
TPI	E-1	Evils in Education (ST)	2.89
TPI	E-3	Hard-Nose Teacher (ST)	2.41
TPI	E-3	Hard-Nose Teacher (OJ)	-2.23
PBI	E-2	Mind vs. Body & Emotions (OJ)	2.20
TPI	E-5	Tough Problem (ST)	1.77**

Table 32

Teacher Evaluation Scale (TES) Regression

Theoretical Factors

Instrument	Factor		t-value
TPOR	T-1	Nature of the Situation	7.02
TPOR	T-7	Motivation and Control	5.17
TPOR	T-4	Use of Subject Matter	3.93
TPI	T-5	Development of Reasoned Hypotheses (OJ)	-3.23
TPOR	T-5	Evaluation	3.10
TPI	T-9	Reliance Upon Extrinsic Motivation (OJ)	-3.07
TPOR	T-2	Nature of the Problem	-2.68
PBI	T-6	Knowing and Doing (ST)	2.47
PBI	T-4	Emotions and Intellect (OJ)	2.35
TPI	T-5	Development of Reasoned Hypotheses (ST)	2.30
TPOR	T-3	Development of Ideas	2.16
PBI	T-2	Change and Certainty	2.11
TPI	T-3	Generation of Ideas	2.10

Table 33

Teacher Evaluation Scale (TES) Regression

Empirical Factors

Instrument	Factor		t-value
TPOR	E-7	Pupil Activity	8.18
TPOR	E-5	Subject Matter Quality	5.32
TPOR	E-6	Generation and Testing of Hypotheses	5.11
PBI	E-3	Knowledge for Its Own Sake (OJ)	-3.49
TPI	E-1	Evils in Education (OJ)	-3.11
PBI	E-3	Knowledge for Its Own Sake (ST)	2.83
PBI	E-5	Religion (ST)	-2.39
TPOR	E-2	Rigidity - Teacher Control	2.21
TPI	E-3	Hard-Nose Teacher (ST)	2.12
PBI	E-5	Religion (OJ)	-1.97

Discussion of the Factor Scores

I. What theoretical factors contribute significantly to variance in observations of the classroom behavior of student teachers as measured by scores on the TPOR?

A. How do beliefs of the student teacher influence the TPOR score received? The student teacher's TPI beliefs on items involving the "Mechanical Following of an Established Method" (TPI Factor T-11) exercised the strongest influence on the observational score. Teachers who said they believed in using lockstep methods got low TPOR scores; those who rejected such beliefs got high TPOR scores. Student teachers who agreed with Dewey that teachers should "Develop Challenging Problems" (TPI Factor T-2) and should involve pupils in "Situations of Experience" (TPI Factor T-1) received high TPOR scores, and vice versa. However, student teachers who agreed with Dewey that teachers should encourage pupils to "Generate Ideas" tended to get low TPOR scores. An easy explanation for this inverse relationship may be that most teachers say that creativity should be encouraged, but that relatively few do anything about it in the classroom.

The only factor in the PBI scores of the student teachers which was found to contribute significantly to the TPOR score consisted of items relating to "Emotions and Intellect" (PBI Factor T-4), and this was inverse. Could it be that these four items accounted for the confusing inverse relationship found between the TPOR score and the total PBI score in the earlier analysis? In any event, the other five PBI factors, which comprise ninety percent of the total PBI score failed to show the significant inverse relationship we got using the total score. This, of course, supports and more fully explains the "hidden power" of the PBI score.

B. How do beliefs of the observer-judge influence the TPOR score given? Observer-judges who were in agreement with Dewey on PBI beliefs involving the relationship of "Knowing and Doing" (PBI Factor T-6), "Science and Morals" (PBI Factor T-3) tended to see teaching which is contrary to that advocated by Dewey, and vice versa. This inverse relationship to the TPOR score is consistent with the findings which resulted from the analysis of the total PBI scores of the observer-judges. Breaking the total PBI scores into theoretical factors, however, did yield one new finding: beliefs involving "Change and Certainty" (PBI Factor T-2) exercise a direct or positive relationship on the observational (TPOR) score. Apparently, relativistic observer-judges are more tolerant of what they are willing to call experimental teaching than are observer-judges who share a broader agreement with Dewey's philosophy. This may also indicate that relativism is the easiest aspect of Dewey's experimentalism to swallow, and is shared by some observer-judges who are not otherwise of an experimental mind.

Clearly, this examination shows that PBI factors exert a stronger influence on TPOR scores given by the observer-judges than do TPI factors. Only items relating to "Reliance Upon Extrinsic Motivation" (TPI Factor T-9) contribute significantly to the influence of the TPI on the observational score.

II. What empirical factors contribute significantly to variance in observations of the classroom behavior of student teachers as measured by scores on the TPOR?

A. How do beliefs of the student teacher influence the TPOR score received? Student teachers who agree with John Dewey by rejecting the sixteen items in the "Evils in Education" factor (TPI Factor E-1), and by also rejecting the three items characterizing the "Hard-Nose Teacher" (TPI Factor E-3) tend to receive high TPOR scores. Teachers who accept these same beliefs tend to receive low TPOR scores. Teachers who agree with Dewey by accepting the two statements making up the "Tough Problem" factor (TPI Factor E-5) are inclined to be seen as experimental, i.e., receive high TPOR scores. These findings indicate that restrictive, corrective, or police-like tactics are not necessary in order to engage pupils with substantial, challenging, or tough problems capable of stimulating thought. These data also seem to indicate that it is the educational beliefs (TPI score) of the student teachers, rather than the philosophical beliefs (PBI score), which are most predictive of how they will be observed to teach (TPOR score).

B. How do beliefs of the observer-judge influence the TPOR score given? The more the observer-judge agrees with Dewey on the "Nature of Learning" (PBI Factor E-6) and on the "Science and Morality" items (PBI Factor E-1) the more likely is he to give out low TPOR scores. Once again the inverse relationship between the observer-judge's PBI score and the TPOR score is demonstrated, virtually duplicating the results obtained in the foregoing analysis of theoretical factors. However, in the present analysis, we did find a direct relationship between the TPOR and the "Mind vs. Body and Emotions" factor (PBI Factor E-2), showing again that an inverse PBI-TPOR relationship does not apply across the board.

Incidentally, comparison of the magnitude of t-values indicates that the empirical factors of the PBI and TPI were more powerful predictors of TPOR scores than were the original theoretical factors.

III. What theoretical factors contribute significantly to the variance in observer-judge evaluations (TES scores) regarding the quality of teaching?

A. How do TPOR factors relate to evaluation of teaching (TES score)? Five of the seven theoretical factors of the TPOR relate directly and positively to the TES score evaluations of teacher behavior. One factor did not turn out to be significant at all, and one factor, "Nature of the Problem" (TPOR Factor T-2) showed an inverse relationship with evaluation. Teachers who otherwise agree with Dewey's experimentalism tend not to agree with him regarding the nature of problems to be dealt with by pupils in school. And even the most experimental of teachers rarely organize classroom activities around problems of the nature advocated by Dewey. Therefore, it is not surprising that this factor shows an inverse relationship with the TES score in this analysis.

Even so, these data corroborate our previous findings that the TPOR score is the most powerful single predictor of the evaluation score given the teacher.

B. How do belief factors of the observer-judges influence the evaluations (TES scores) given the student teacher? The more strongly the observer-judge believes in the "Development of Reasoned Hypotheses" (TPI Factor T-5), the more likely he is to give student teachers a lower evaluation. Likewise, he tends to give a low evaluation score if he agrees with Dewey in condemning "Reliance Upon Extrinsic Motivation" (TPI Factor T-9). Apparently these two factors make the strongest contribution to the general inverse relationship between educational beliefs (TPI) and evaluation of teacher competence (TES) found in the earlier analysis using total TPI scores.

Whereas PBI Factor T-4, "Emotions and Intellect," had an inverse effect on the TPOR score, it has a direct effect on the evaluation (TES) score. This phenomenon seems to indicate that observer-judges who believe in the continuity of emotions and intellect see very little experimental teaching in classrooms, but that they tend to like what they see. It could be that the items in the "Emotions and Intellect" factor of the PBI attract people who are not otherwise in agreement with experimentalism, increasing the chances that non-experimental teaching will be given higher evaluations than if this factor attracted only thoroughgoing experimentalist observer-judges.

C. How do belief factors of the student teachers influence the evaluations (TES scores) they receive? Student teachers who have high scores on PBI Factor T-6, "Knowing and Doing," and on TPI Factor T-5, "Development of Reasoned Hypotheses," are likely to be given high evaluation scores, and vice versa. It is interesting to recall that observer-judges who hold high scores on the "Development of Reasoned Hypotheses" factor tend to give low evaluation scores. Apparently when they do encounter a student teacher who shares their experimental views on this factor they reward that teacher with a high evaluation.

IV. What empirical factors contribute significantly to the variance in observer-judge evaluations (TES scores) regarding the quality of teaching?

A. How do TPOR factors relate to evaluation of teaching (TES score)? Again, TPOR factors lead the way in influencing the evaluation score, directly and positively. Teachers who were seen to provide for a great deal of "Pupil Activity" in the classroom were given higher ratings (TES scores) than teachers who did not. Likewise, teachers who provided "Subject Matter Quality" of a challenging nature which went beyond regurgitation of textbook answers were given higher ratings than teachers who did not. Teachers who engaged pupils in activities calculated to "Generate and Test Hypotheses," or, in other words, those who were seen to teach in the "hypothetical mode," were rated as better teachers than those who did not. Teachers who refrained from exercising rigid and tight control of classroom activity got better ratings than those whose behavior was characterized by "Rigidity and Teacher Control."

B. How do belief factors of the observer-judges influence the evaluations (TES scores) given the student teachers? Observer-judges who agreed with Dewey in his rejection of the notion of "Knowledge for Its Own Sake" (PBI Factor E-3) tended not to like the teaching they observed, presumably because they saw so much teaching which emphasized the acquisition of knowledge and skills as an end in itself. Non-experimental observer-judges who believed in "Knowledge for Its Own Sake" tended to give favorable ratings, presumably because they saw plenty of evidence of this kind of teaching.

Observer-judges who agreed with Dewey in the rejection of conventional religious beliefs (PBI Factor E-5) tended to give lower ratings, and vice versa.

The large "Evils in Education" factor (TPI Factor E-1) had a pronounced inverse relationship on the evaluation scores. Observer-judges who accepted these "evils" tended to like what they saw, those who rejected them did not. Presumably, there were plenty of these "evils" in evidence.

C. How do belief factors of the student teachers influence the evaluations (TES scores) they receive? Student teachers who agreed with Dewey in his rejection of "Knowledge for Its Own Sake" (PBI Factor E-3) received higher evaluations than those who disagreed with Dewey on this score. We should recall that observer-judges who held high scores (agreement with Dewey's experimentalism) on this same factor tended to give low evaluations. Therefore, we may conclude that when they did encounter a student teacher who shared their views about "Knowledge for Its Own Sake" they rewarded that teacher with a favorable evaluation.

A similar comparison can be made on the "Religion" factor (PBI Factor E-5). Student teachers who held conventional views on religion received high evaluations, and those who shared Dewey's experimental views on religion tended to receive low evaluations. It needs to be pointed out that the majority of observer-judges held conventional or non-experimental religious beliefs, as did the majority of student teachers. However, a substantial minority of the observer-judges strongly opposed traditional religious beliefs, and this group was most critical of the teaching they observed--regardless of the views of the student teachers on religion. They did not seem to reward (or recognize?) student teachers who shared their religious beliefs.

Experimental student teachers who rejected beliefs that teachers should be "hard-nosed" (TPI Factor E-3) tended to get better ratings than those who believed the teacher should maintain strict disciplinary control.

Summary of Factor Score Analysis

By breaking scores of the belief and observation instruments into theoretical and empirical factors it has been possible to more fully explore the interrelationships between behavior of observer-judges and the student teachers they observed and evaluated. The examination of factors has more clearly explicated the inverse relationships of the Personal Beliefs Inventory and the Teacher Practices Observation Record as well as identified the more powerful factors which influence both observation scores and evaluations.

The factor analysis and multiple linear regression procedures have shared in elucidating some of the very complex interactions between beliefs, observations and evaluations. Also they have clearly identified some classroom practices which nearly all observer-judges see as good and reward with high evaluations. The teacher who focuses classroom attention upon the pupils and their activities, who grants students opportunities for freedom of expression, who develops lessons which go beyond the simple acquisition of material presented in textbooks and who encourages self-discipline and internal motivation for learning is the teacher that observer-judges will evaluate highly. Most observer-judges, although of varying occupations and belief systems, value these teacher behaviors as good when they see them.

CHAPTER V

PHASE III (FOLLOW-UP STUDY)

Purposes

The objectives of Phase III were to:

1. Follow up the student-teacher subjects of Phase II during their first year of service as certified teachers through continued observations and evaluations by observer-judges.

2. Compare the observations and evaluations of the classroom behavior of first-year teachers with those of experienced teachers.

3. Compare the belief scores, classroom observation scores, and evaluations of the subjects as student teachers with those of the subjects as first-year teachers.

Procedures

Selection of the Sample. From the total of 407 student teachers who had served as observer-subjects from Phase II, 100 subjects were chosen for study in Phase III. The 100 subjects were selected on the basis of (1) their status as a certified, employed teacher, and (2) the completeness of the data collected on them from Phase II. In addition, 100 experienced teachers[1] were added to the sample. Each of the experienced teachers was selected randomly from the faculty of the school in which the first-year teacher was serving. Thus, each first-year teacher was matched with an experienced teacher, chosen randomly, who was employed in the same school and who taught at the same grade level or in the same general subject matter area.

The 300 observer-judges of Phase III were selected from the same school systems as the observees. This group consisted of principals, supervisors, central office pesonnel, and classroom teachers. Each pair of teachers, first-year and experienced, were repeatedly observed and evaluated by (1) the principal of the school in which they served, (2) another principal serving in the same district, and (3) a member of the supervisory or teaching staff of the school system. Thus, each pair of subjects were repeatedly observed by three observer-judges. In addition, three observer-judges from Phase II, university personnel, also observed and judged a portion of the Phase III subjects.

---

[1] Mean years of experience was 6.97

Altogether 303 observer-judges observed and evaluated 200 classroom teachers for a total of 1,892 observations and ratings.

Data Collection

Every subject (teacher and observer-judge) completed the *Study of Beliefs* (including the PBI, TPI, and the D-Scale) prior to the observation period. Three observer-judges individually observed and evaluated each pair of teachers three times, using the *Teacher Practices Observation Record* (TPOR) and the *Teacher Evaluation Scale* (TES). Observations were scheduled for the first three months of 1967; each teacher was visited by the three observer-judges in January, again in February and finally in March. In total, nine observations and evaluations of competence were made for each teacher-subject.

Analysis of Data

The scores given the teachers by the observer-judges on the TPOR and TES were submitted to multiple regression analysis in order to identify variables which contributed to the variance of observation and evaluation scores for both first-year and experienced teachers. Each of these two scores were treated separately, and in turn, in a series of increasingly complex regression models.

Teacher Practices Observation Record. Four models were proposed in order to determine which variables contribute most to the TPOR scores, which served as the response.

Model 1: This general model was fitted with 21 variables:

PBI TPI POQ	observer-judge	3 vars.
PBI TPI POQ	teacher	3 vars.
15 two-way interactions of PBI, TPI, and POQ of teacher and observer		15 vars.

The F-value was 7.00 which is significant at the .01 level. This regression model contributed a signficant amount of information about the TPOR scores.

Model 2: This general model was fitted with 40 variables. The personal characteristics of observer and teacher were brought into the model as main effects, their interactions were not included. The 21 variables from Model 1 were included with the personal characteristic variables to produce the following:

Observer Occupation	4 vars.
Time of Observation	2 vars.
Institution	5 vars.
Length of Service of Teacher	1 var.
Teaching Level	1 var.
Subject Matter Taught	6 vars.
Variables from Model 1	21 vars.

The calculated F-value was 9.17, again significant at the .01 level. Thus, the model accounts for a significant amount of response variance.

Model 3: Model 3 was a 39-variable model which was designed to identify relationships between teacher and observer characteristics as they interact with belief scores in accounting for variance of the Teacher Practices Observation Record scores. The list of variables follows:

Date of Observation	2 vars.
Length of Service of Teacher	1 var.
Teaching Level	1 var.
Beliefs of Observer-judge	3 vars.
Beliefs of Teacher	3 vars.
Date x Occupation of Observers	2 vars.
Date x Teaching Level	2 vars.
Occupation of Observers x Teaching Level	1 var.
Date x Observers' Beliefs	6 vars.
Date x Teachers' Beliefs	6 vars.
Occupation of Observers x Observers' Beliefs	3 vars.
Occupation of Observers x Teachers' Beliefs	3 vars.
Teaching Level x Observers' Beliefs	3 vars.
Teaching Level x Teachers' Beliefs	3 vars.

The F-value of the model was 4.06, which indicates this to be significant at the .01 level; thus the variables make a significant contribution in accounting for response variance.

Model 4: This final model was designed to explore the relationships of the belief scores of the observer-judge and the teacher with the subject matter taught in accounting for variance of the TPOR. The variables were as follows:

Subject Matter	6 vars.
Observers' Beliefs	3 vars.
Teachers' Beliefs	3 vars.
Subject Matter x Beliefs	36 vars.

This model had an F-value of 4.39, significant at the .01 level.

Findings

I. Variables which contribute significantly to the variance in observations of the classroom behavior of teachers as measured by the scores on the Teacher Practices Observation Record are shown in Table 34.

II. Variables which contribute significantly to the variance in evaluations of teacher competence as measured by TES scores are shown in Table 35.

Discussion

I. What variables and combinations of variables contribute significantly to the variance in observations of the classroom behavior of teachers as measured by scores on the TPOR?

A. How do the beliefs of the teacher influence the TPOR score received? Model 1 indicates that the teacher's PBI score has a slightly positive effect on the teacher's observed classroom behavior (TPOR score) if his D-Scale score is high (open-minded) and the observer-judge's D-Scale score is low (closed-minded). However, as the teacher's D-Scale score decreases and the observer-judge's increases, the effect of the teacher's PBI on the TPOR becomes negative in nature.

Model 1 also indicates that the teacher's TPI score is directly related to the TPOR score given the teacher's observed teaching behavior. This relationship is even more pronounced if the teacher is also in high agreement with experimentalism on the PBI.

These data pretty much corroborate the findings in Phase II (pp. 59 and 67). The teacher's beliefs about specific classroom practices (TPI score) are more predictive of the teacher's observed classroom behavior (TPOR score) than are beliefs about the fundamental questions of philosophy (PBI score).

In general, the teacher's D-Scale score and the TPOR are positively correlated, with this relationship being more pronounced if the PBI is high and the TPI low, but becomes modified as these scores reverse themselves. Again, the PBI seems to provide "extra kick" to the influence of the D-Scale on the TPOR, just as it does for the influence of the TPI on the TPOR.

B. How do beliefs of the observer-judge influence the TPOR score given? Again (as in Phase II, p. 59) it was found from Model 1 that the observer-judge PBI has slightly negative effect on the TPOR. However, this is the case only when the observer-judge TPI score is low. As the observer-judge TPI score increases the correlation becomes increasingly positive.

The effect of the observer-judge's TPI is generally positive, with this effect very pronounced if his PBI and D-Scale scores are high. (Remember, in this study a high D-Scale score indicates open-mindedness.)

Table 34

Variables Which Contribute to Variance In Observation of Classroom Behavior

Response: TPOR Total Scores

	Variable		t-value
Model 1:	R^2 = .0729 (F = 7.00**)		
	T's PBI x T's D-Scale		2.98**
	T's TPI x T's D-Scale		-2.96**
	O's PBI x O's TPI		2.16**
	O's TPI x O's D-Scale		1.79*
	O's D-Scale x T's PBI		-1.69*
	O's PBI x O's D-Scale		-1.47
	O's D-Scale x T's D-Scale		-1.46
Model 2:	R^2 = .1169 (F = 6.12**)		
	Position -	1	1.46
		2	1.09
		3	4.45**
		4	-1.98**
	Institution -	1	2.12**
		2	-2.54**
		3	3.85**
		4	-1.58
		5	3.35**
	Grade Level -		-3.38**
	Subject -	1	0.22
		2	-0.06
		3	-1.73*
		4	1.16
		5	-2.41**
		6	-2.49**
	T's TPI x T's D-Scale		-3.05**
	T's PBI x T's D-Scale		2.41**
	O's D-Scale x T's PBI		-2.32**
	O's TPI x O's D-Scale		2.13**
	O's PBI x O's TPI		1.36
Model 3:	R^2 = .0787 (F = 4.06**)		
	Teaching Level x T's PBI		-3.41**
	Teaching Level x O's D-Scale		2.91**
	T's Experience (New or Experienced) x T's PBI		-1.74*
	T's Experience x T's D-Scale		-1.54
	O's TPI		3.72**
	T's TPI		3.01**
	O's PBI		-2.06**

** Significant at the .05 level
* Significant at the .10 level

Note: T's = Teachers or Observers
O's = Observers

Table 35

Variables Which Contribute to Variance In Evaluations of Teaching Competence

Response: TES Total Scores

	Variable	t-value
Model 1:	$R^2 = .2741$ (F = 25.12**)	
	O's D-Scale x T's TPI	-4.20**
	T's PBI x T's D-Scale	-3.73**
	O's D-Scale x TPOR	2.99**
	O's D-Scale x T's PBI	2.51**
	O's TPI x T's PBI	-2.51**
	O's TPI x O's D-Scale	2.23**
	T's TPI x TPOR	2.26**
	O's TPI x T's D-Scale	-1.89*
	O's TPI x T's TPI	1,83*
	O's PBI x T's PBI	-1.63
	T's D-Scale x TPOR	-1.53
Model 2:	$R^2 = .3399$ (F = 20.20**)	
	Position - 1	0.96
	2	-1.67*
	3	1.33
	4	0.89
	Institution - 1	2.14**
	2	4.71**
	3	1.66*
	4	-2.81**
	5	-3.03**
	T's Position -	-8.71**
	Subject - 1	-1.23
	2	0.93
	3	-2.95**
	4	-3.41**
	5	5.33**
	6	-2.01**
	O's D-Scale x T's PBI	-4.22**
	T's PBI x T's D-Scale	-3.32**
	O's D-Scale x TPOR	2.90**
	O's TPI x O's D-Scale	2.79**
	O's D-Scale x T's PBI	2.46**
	O's TPI x T's PBI	-2.14**
	O's TPI x T's D-Scale	-1.74*
	O's PBI x T's TPI	1.70*
	T's TPI x T's D-Scale	-1.66*
	T's D-Scale x TPOR	-1.41
	T's TPI x TPOR	1.37

Table 35 (continued)

	Variable	t-value
Model 3:	$R^2 = .3021$ (F = 18.17**)	
	Teaching Level x TPOR	3.23**
	Teaching Level x T's TPI	2.55**
	T's Position x Teaching Level	-1.98**
	T's Position x T's PBI	1.94*
	Date of Observation x Teaching Level - 1	1.68*
	2	0.31*

**Significant at .05 level
*Significant at .01 level

T's = Teachers or observers
O's = Observers

The effect on the TPOR of the observer-judge's score on the D-Scale seems in general to be slightly negative. If both the PBI and D-Scale scores are high for the teacher being observed, this effect becomes very signficant. On the other hand, as the teacher's PBI and D-Scale scores are modified, this effect may become slightly positive, particularly as the observer-judge's TPI increases.

C. _What factors other than beliefs of the observer-judge and the teacher influence TPOR scores_? Model 2 shows that the occupational position of the observer-judge made a significant contribution in accounting for variance of the TPOR score. Public school central office personnel tended to give the highest scores, followed in order by supervisors, principals, and teacher-observers. University observers were most stingy in giving out high TPOR scores.

The time of year or date of the observation made no significant contribution to TPOR variance.

The teacher education institution variables did, however, contribute significantly. The observations made at Northwestern University yielded the highest TPOR scores, followed by Oneonta, Sacramento, and Wisconsin. The lowest TPOR scores were given at Albany and Florida. The reader must be cautioned not to interpret these findings as indicating that the teaching produced by Northwestern is more experimental than that produced by Florida. The experimentalism of the observer-judges is a major factor in determining TPOR scores, and these data may simply indicate that Florida observer-judges held higher expectations with respect to experimentalism than did the Northwestern observer-judges. The purpose of this study was not to compare our cooperating institutions, and if one were to attempt to do so a much more thorough examination of these data toward this end would be required.

Whether the teacher being observed was "new" or "experienced" made no significant contribution to variance on the TPOR scores within the framework of the analysis used in Model 2. A more detailed report on the comparison of "new" and "experienced" teachers is given later in this chapter.

A difference was found, however, between secondary school teachers and elementary school teachers. Secondary teachers received significantly higher TPOR scores than those at the elementary level. Again, one should not leap to the conclusion that elementary teachers are less experimental than are their secondary school counterparts. The beliefs and expectations of the observer-judges at these respective levels plays an important role in determining TPOR scores.

The subject being taught contributed slightly to the variance in the TPOR score received. Reading teachers (for the most part at the elementary school level) received the highest TPOR scores, while mathematics teachers (at both the secondary and elementary levels) and foreign language teachers received the lowest TPOR scores. The other subject areas fell between these extremes.

Model 3 indicates that the level of teaching interacts significantly with the observer-judge's D-Scale score. The effect of the observer-judge's D-Scale score is not significant when elementary teachers are being observed, but has a significantly negative effect on the TPOR if the teaching is on the secondary school level. This means that a closed-minded observer-judge at the secondary level is inclined to see lots of experimental teaching, and vice versa.

Model 3 also indicates that the experience of the teacher interacted significantly with his PBI and D-Scale score. The effect of the PBI was positive for both new and experienced secondary teachers. In the case of elementary teachers, the effect of their PBI is negligible for experienced teachers, but significantly negative for new ones.

II. What variables and combinations of variables contribute significantly to the variance in judgments made with respect to the quality of teaching which was observed?

A. How does the observation score (TPOR) relate to judgments of teacher competence (TES score)? Model 1 shows that the relationship between the TPOR and the TES is direct if the observer's D-Scale score and the teacher's TPI are average to high, but the effect becomes negligible when these scores are low. In other words, an open-minded observer-judge who observes a teacher whose beliefs are in agreement with experimentalism will tend to see experimental teaching behavior and to reward that behavior with high evaluative ratings. However, if the judge is close-minded and the teacher is fairly ambivalent about experimentalism, the observer-judge may give the teacher a high TPOR score but will not necessarily support it with a favorable TES rating.

Model 3 shows that the TPOR score, as in Phase II, was found to be the most significant factor influencing the observer-judge's TES rating of teacher competence. The effect of the TPOR on the evaluation scores was again a very pronounced positive one, with this effect being even more pronounced if the teaching took place at the elementary school level.

Observer-judges in general give their highest ratings to teaching they see to be in agreement with Dewey's experimentalism, and vice versa.

B. How do the beliefs of the observer-judges influence their judgments of teacher competence? Model 1 indicates that once again (see p. 61) the observer-judge's PBI on the TES rating is inverse. This effect is most pronounced when the teacher being evaluated has a high PBI and a low TPI. However, if the teacher has a low PBI and a high TPI, the effect of the observer-judge's PBI on the evaluation becomes negligible.

The effect of the observer-judge's TPI on the TES rating is also generally negative. This effect is most pronounced when the teacher's PBI and D-Scale score are high and when the teacher's TPI and the observer-judge's D-Scale score are low. As the former scores drop and the latter ones increase, the effect becomes less significant. In short, observer-judges who do not agree with Dewey's experimentalism and who are also rather dogmatic in their beliefs tend to give teachers higher ratings than do observer-judges who are both experimental and open-minded.

The observer-judge's D-Scale continues the generally inverse trend of the effect of beliefs on his evaluations of the teacher. This effect is most negative when the observer has recorded a low TPOR score for a teacher who has a low score on both PBI and TPI. Clearly, the open-minded observer-judge, like the high PBI-high TPI observer-judge, expects "good" teachers to perform experimentally.

C. How do the beliefs of the teachers influence the ratings given their observed teaching behavior? As has been the case in earlier analyses, the effect of the teacher's PBI is not clear-cut. The effect of the teacher's PBI on his rating is positive providing his TPI is high and both the observer's PBI and TPI are low. However, this effect seems to be due more to the generally inverse effect of the observer's belief scores than to the PBI of the teacher.

The nature of the relationship of the teacher's TPI to the evaluation score is also determined by other belief scores in the picture. If the teacher's PBI and TPOR scores are also high, his TPI enjoys a positive relationship to the rating he receives.

The relationship between the teacher's D-Scale score and the evaluation (TES score) he receives reaches each extreme as other factors change. If the observer's TPI is low, his D-Scale score is high, and the teacher's PBI is low, then the relationship is significantly positive. But if the observer's TPI is high, his D-Scale score low, and the teacher's PBI is high, the relationship becomes very significantly inverse in nature.

D. What other descriptive factors influence ratings of teacher competence? Model 2 indicates that the occupational position of the observer-judge has no apparent effect on the evaluations (TES scores). Likewise, the date on which the observation was made (i.e., whether it was the first, second, or third observation) seems to be insignificant.

However, the institution at which the evaluation was made has some effect. The Phase III teachers associated with Albany were given the highest evaluations, followed in order by Sacramento, Oneonta, and Florida, with Northwestern and Wisconsin lowest. This, of course, does not necessarily mean that Albany produces the "best" teachers. It simply means that all of the factors measured in this study interacted in such a manner that observer-judges who evaluated Albany graduates (and the experienced teachers paired with them) could be counted on to award higher ratings than those who evaluated teachers from the other institutions.

Model 2 also shows very clearly that experienced teachers were given much higher ratings than were the new teachers followed up from Phase II. The majority of observer-judges in Phase III were public school principals and supervisors, and this could reflect a bias on the part of education practitioners for experience, particularly since we found no significant difference on the TPOR scores between "old" and "new" teachers.

Whether the subjects taught at the elementary or secondary level did not contribute significantly to variance in the TES evaluations. It is interesting to note, however, that foreign language teachers received the highest evaluations. The lowest evaluations were given to teachers of mathematics, English, speech, journalism and language arts. Teachers of reading, natural science, and social science fell between the extremes.

Although Model 3 accounted for about 30 percent of the variance in the evaluations of teaching competence (TES scores), very few interactions in this model proved significant. This may be due to the fact that the TPOR score based on systematic classroom observations so clearly overshadowed all other variables and combination of variables in accounting for variance in ratings of teachers.

However, Model 3 does provide several important insights not provided in earlier analyses of the data. For example, contrary to the general trend favoring experienced teachers over new teachers, new elementary school teachers received significantly higher evaluations than did experienced elementary teachers.

The effect of the teacher's PBI on evaluation scores was affected somewhat by the teacher's position. The PBI's effect was most negative for a new elementary school teacher, and becomes slightly positive for an experienced teacher. Apparently, the basic experimentalism of teachers "shows" more clearly and is valued more highly in seasoned high school teachers than among inexperienced grade school teachers.

On the other hand, the teacher's TPI had a most positive effect on the evaluation if the teacher was experienced and taught in the elementary school. This effect was greatly modified for new secondary teachers. This means high agreement with Dewey's educational philosophy was more essential to high ratings for "old" elementary teachers than for "new" high school teachers.

Model 3 shows that secondary teachers generally receive significantly higher ratings than do elementary teachers, but that this effect is reversed if the teacher's TPI is high, or if the TPOR received is high, and if the teacher is new. High school teachers seem to be able to get high ratings without necessarily agreeing with Dewey's experimentalism either in theory or in practice, but elementary teachers cannot, particularly if they are "old-timers." Or to put it still another way, observer-judges at the elementary school level clearly pay off on experimentalism. In contrast, evaluations at the secondary level are not nearly so sensitive to the experimentalism factor.

Conclusions

The Phase III study generally corroborated the results yielded by Phase II. However, the comparisons of experienced and new teachers permitted by the design changes for Phase III yielded several very interesting findings, particularly with respect to evaluation scores. The major reason for the addition of experienced teachers to the Phase III study was the predominantly high evaluation scores given the student teachers in Phase II. It was hoped that we might find a larger spread of "good" and "bad" teachers; instead, we just found more "good" teachers with even higher evaluation scores. Even so, Phase III provided a number of new insights into what happens when observer-judges are asked to rate teacher competence on the basis of systematic observations of classroom behavior. A comparison analysis of Phase II and Phase III data is presented in Chapter VI.

CHAPTER VI

COMPARISON ANALYSIS OF PHASE II AND PHASE III DATA

There were 84 subjects who participated as student teachers in Phase II and then one year later as new teachers in Phase III. Actually, 100 subjects participated in both Phase II and Phase III, but observations were complete for only 84 of these. It is the data on these subjects which served as the basis for the analysis described in this chapter.

The first step in the comparison analysis was a simple paired difference t-test on the various instruments administered in both Phase II and Phase III. The results are presented in Table 36.

Note that the observer-judge PBI decreased significantly from Phase II to Phase III. This reflects the shift from college-controlled to public school-controlled appointment and selection of observer-judges. This corroborates our earlier findings that public school people are much less enthusiastic supporters of Dewey's basic philosophy than are college professors.[1] This, according to the clearest finding in both Phase II and Phase III, would lead one to expect the significant increase in TPOR shown in Table 36. The fact that Phase III observers were less experimental in their basic beliefs tends to make teaching practices appear more experimental to them. Likewise, the observer-judge who does not believe in the fundamental tenets of experimentalism is likely to give an unfavorable opinion of the seemingly experimental teaching practices he "sees," which is reflected in the lower TES ratings for our 84 subjects in Phase III, though the difference is not statistically significant.

The teachers in Phase III made higher scores on all three beliefs instruments than they made as student teachers in Phase II, though only the D-Scale increase was statistically significant. It is doubtful that this is due to growth during the intervening year. It is more likely due to the fact that the new teachers responded to identical forms of the PBI, TPI, and D-Scale the previous year when they were student teachers. However, it is important to note that the belief gap widened considerably between observer-judges and teachers as the study moved from the college scene to the public school scene, particularly at the fundamental level represented by PBI and D-Scale scores. It should also be noted that as new teachers take their initial teaching positions in the public schools they move into situations in which there is increased discrepancy or conflict between basic philosophical beliefs and educational beliefs. Notice that Phase III observer-judges were slightly more experimental on the TPI (educational beliefs) but significantly less experimental on the PBI. These data provide some insight into the impact of the public schools on the belief systems of beginning teachers.

---

[1] Bob Burton Brown, The Experimental Mind in Education (New York: Harper and Row, 1968), Part III.

Table 36

Results of Paired Difference t-tests Comparing Phases II and III

Instrument	Differences in Means (III - II)	S.D.	t-value
Observer-Judge's			
PBI	-4.616	17.642	-2.398*
TPI	2.297	20.429	1.030
D-Scale	-2.184	15.812	-1.266
Teacher's			
PBI	2.060	15.838	1.192
TPI	0.738	16.207	0.417
D-Scale	4.690	14.999	2.866*
TPOR	5.654	16.482	3.144*
TES	-0.550	4.795	-1.050

* Significant at .05 level.

Perhaps the most important finding reported in Table 36 is that there was no significant difference in the ratings of teaching competence (TES scores) from Phase II to Phase III. While the Phase III judges didn't like the new teachers quite as much as did the Phase II judges, one must conclude that evaluations made of the student teachers held up pretty well a year later in the public schools.

Since there was no significant difference in the TES ratings further analysis of these ratings seemed useless. Therefore, we concentrated our further study on the significant increase in the TPOR observational score. Several regression models were proposed using the difference between the mean TPOR score in Phase III and the mean TPOR score in Phase II as the response.

Model 1 (Table 37) accounted for an amazing 78 percent of Phase II-Phase III variation in the TPOR. By far the most significant predictor was the Phase II TPOR. An inverse relationship exists, indicating that if the student teacher's Phase II TPOR was high, the Phase III TPOR did not vary much from it; but if the Phase II TPOR was low, there tended to be a significant increase in the Phase III TPOR. This finding strongly suggests that the Phase III observer-judges were not very discriminating in recording high TPOR scores, or at least not so discriminating as their counterparts in Phase II.

The observer-judge's TPI seemed to play a significant role in the prediction of TPOR change. A positive relationship between Phase II and III TPI was found with the difference in Phase II-Phase III TPOR scores. This would seem to indicate that if the Phase II observer-judge TPI scores were high, the TPOR would tend to increase. And if the Phase III observer-judges had higher TPI scores than the Phase II observer-judges, this too would tend to increase the TPOR scores. Or, in other words, observer-judges who wanted to see experimental teaching practices tended to see them, with this tendency accentuated among Phase III (public school) observer-judges.

Model 2 (Table 38) employed 9 variables, and was designed to screen out important personal characteristics in predicting TPOR change. Factors studied included the teacher training institution at which the observations took place (3 vars.), the sex of the teacher, the educational level (elementary or secondary) at which the teaching was done in Phases II and III (2 vars.) and the interaction of these two "level" variables. Institution seemed to play a significant role in determining the size of TPOR change, with the greatest positive change occurring at Oneonta, the least positive change (slightly negative in nature) occurring at Florida. The sex of the teacher was very close to being significant, with the female teacher receiving a greater increase in TPOR score than the male teacher.

Finally, a third model was hypothesized in order to explore the possibility of interactions of variables in Models 1 and 2. Very little additional significance was found, although an interesting interaction between sex and the Phase II TPOR occurs. The inverse relationship between the Phase II TPOR and subsequent TPOR increase is much more pronounced for female teachers than for male teachers.

Table 37

Multiple Regression - Model 1
Response: Difference in Mean TPOR Scores

$R^2 = 0.7841$

Variable			t-value
PHASE II			
	Observer-Judge	PBI	0.07
	"	TPI	2.04*
	"	D-Scale	-0.66
	Student Teacher	PBI	1.00
	"	TPI	0.98
	"	D-Scale	0.43
		TPOR	-10.22*
		TES	1.09
PHASE III			
	Observer-Judge	PBI	0.32
	"	TPI	2.32*
	"	D-Scale	0.14
	Teacher	PBI	-0.50
	"	TPI	1.07
	"	D-Scale	0.58
		TPOR	--
		TES	0.65

F = 14.10 *

* Significant at .05 level

(Note: The t-value is a result of testing the variable's coefficient for significant difference from 0. The F is a result of analysis of variance to determine whether the model accounts for a significant amount of the response variance.)

Table 38

Multiple Regression - Model 2

$R^2 = .2130$

Variable	t-value
Institution Varibles	
1 - Sacramento	0.24
2 - Albany	0.19
3 - Oneonta	2.27*
4 - Wisconsin	-0.40
5 - Northwestern	0.62
Observee's Sex	-1.58
Level (Elem. or Sec.) of Response in Phase II	0.18
Level of Response in Phase III	0.28
Level x Level	-1.33

F = 2.23*

*Significant at .05 level

The Phase II TPOR also seemed to interact somewhat with the Phase II-III change in the teacer's TPI. The TPOR's negative relationship with the Phase II TPOR is modified if the teacher's TPI difference is positive and large, but increases if the difference is negative. This is simply futher evidence of the high correlation between TPOR score and the teacher's TPI. The teacher who falls away from experimentalism by lowering his TPI beliefs may also pull down lower TPOR scores.

In summary, it should be remembered that by far the most significant predictor of the change in TPOR score from Phase II to Phase III was the Phase II TPOR in a simple linear relationship. Or, in more understandable terms, the student teacher who was recorded as teaching in high agreement with Dewey's experimentalism can be counted on to be observed teaching that way (and even more so) a year later as a beginning teacher in the public school. TPOR scores are, indeed, reliable predictors of future TPOR scores, which says absolutely nothing in view of our findings regarding the importance of knowing who took the TPOR observations and what his beliefs were.

CHAPTER VII

SUMMARY, CONCLUSION, AND RECOMMENDATIONS

The study of observer-judges' ratings of teacher competence departed from the usual pattern of research on teacher competence. The observer-judges were studied much more pointedly than were the teachers. The basic purpose of the study was to find out what happens when persons are asked to make evaluations of teacher competence based on structured observations of teaching performance. It definitely was not the purpose of the study to come up with final answers to the question "Who is a good teacher?" Instead, the focus was placed on discovering some factors which must be dealt with if we are to find answers to such a question which could possibly make sense. If anything, the study sought answers to the question "Who is a good observer-judge of teacher competence?" Or, better still, "What is a good procedure for making observations and evaluations of teacher competence?"

In order to study the observer-judge and his use of the instruments for observation and evaluation, it was essential that no attempt should be made to "train out" the differences which might be found among them. Therefore, the observer-judges were simply taken "as they came" from professional groups normally associated with teacher training, and they were provided only the most cursory sort of instructions with respect to the use of the instruments. They received no training designed to push up artificially the reliability and validity of their observations and evaluations. Instead, assessments of their beliefs were taken in an attempt to account for or "figure in" a possible source of unreliability and invalidity in their observations and evaluations of teacher competence. This element in the basic design contributed to the major findings of the study.

The study of observer-judges' ratings of teacher competence involved the collection of a vast amount of data over a period of four years. Multiple and repeated observations and evaluations were made by observer-judges of teachers' classroom performances. In all, 972 observer-judges made 5,201 observations and 3,295 evaluations of 612 teachers in grades 1 to 12 in widely varying schools across the nation.

Student teachers whose pre-service teaching performances were evaluated were drawn from those students enrolled in teacher education programs at Sacramento State College, New York State University at both Albany and Oneonta, Northwestern University, the University of Wisconsin, and the University of Florida. The observer-judge teams for the student teachers were selected from the faculties of these colleges and the public schools which cooperated in their teacher education programs. One hundred of the student teachers were followed to their first in-service teaching experience the next year, and were compared with an equal number of experienced teachers. Only a small number of observer-judges from the earlier phase could be used in the follow-up study, so public school supervisory personnel then assumed the observer-judge responsibility.

Instruments used in the collection of data included the Personal Beliefs Inventory (PBI), which measures agreement-disagreement with Dewey's experimentalism at the level of fundamental philosophic beliefs; the Teacher Practices Inventory (TPI), which measures agreement-disagreement with Dewey's experimentalism at the level of beliefs about educational practices, and the Dogmatism Scale (D-Scale), which measures the structure of belief systems along an open-closed dimension. Descriptions of the classroom behavior of teachers was obtained by using the Teacher Practices Observation Record (TPOR), which measures observed teacher practices in terms of Dewey's philosophy of education. Judgments of teacher competence were recorded on the Teacher Evaluation Scale (TES), which is a simplified rating form.

The data were submitted to multiple regression analysis to discover which of the measured variables contributed to scores given teachers by the observer-judges on the Teacher Practices Observation Record (TPOR) and the Teacher Evaluation Scale (TES).

Summary of Findings

A. Phase I involved the observation and evaluation of five filmed teaching episodes by 130 observer-judges at four widely separated teacher education institutions. Phase I provided an opportunity to select, organize, and orient the observer-judges, as well as to field test the instruments and procedures used in the phases of the study which followed.

1. It was possible to compute reliability coefficients for the use of the TPOR by untrained observers. A new concept of "within-observer" reliability was based on the observer's consistency in recording the same (filmed) teaching behavior over a period of time. TPOR observers had "within-observer" reliability coefficients ranging from .48 to .62. Using more conventional techniques for estimating reliability, TPOR observers were found to have "between-observer" reliability of .57. The internal consistency reliability coefficients found for the TPOR ranged from .85 to .93.

2. Reliability for the CBRS (Ryans-type observational record) was also established under conditions identical to those for the TPOR. Within-observer reliability was excellent, ranging from .79 to .86. Likewise, internal consistency ranged from .71 to .85.

3. The observation scores on both the TPOR and the CBRS showed direct relationship with the ratings (TES scores) of teacher competence. The more a teacher was seen to perform in agreement with Dewey's experimentalism (high TPOR scores) and non-authoritarianism (low CBRS scores) the higher that teacher's rating on the Teacher Evaluation Scale (TES).

4. The Classroom Behavior Rating Scale (CBRS) was found to be very poor at discriminating among teachers. Very little of the variance in CBRS scores could be attributed to the diverse teaching styles and personalities of the different teachers observed. The teacher variable accounted for only thirteen percent of the variance in TPOR scores. CBRS scores closely paralleled the TES ratings, leading to the conclusion that they represented evaluative ratings rather than descriptive observational reports. Therefore, the CBRS was not used to collect observational data in Phases II and III.

5. The Teacher Practices Observation Record (TPOR) was found to have substantial power for differentiating teachers. The teacher variable accounted for forty-three percent of the variance in TPOR scores. Therefore, the TPOR was considered adequate for gathering observational data in subsequent phases of the study.

B. Phase II involved making observations and evaluations of 407 student teachers' classroom performances in pre-service clinical teaching situations by 539 observer-judges. The observer-judges included the college and public school personnel which normally observes and evaluates student teachers plus an added number of education and academic professors who were induced to participate in this particular study. The major findings of this phase were as follows:

I. Observations of Student Teacher Behavior

a. Student teachers whose professed educational beliefs are in strong agreement with Dewey were seen by observers as practicing many of the teaching behaviors advocated by Dewey. If in addition, their personal philosophic beliefs are also in agreement with Dewey, their teaching behavior was seen as even more Dewey-like, provided the observer also held beliefs in high agreement with Dewey's experimentalism. If a student teacher was to be seen as successful in putting his experimental educational beliefs into practice, it seemed necessary that he (1) also hold basic philosophical beliefs which gave congruent support to his professed educational beliefs, and (2) be observed by an individual who holds these same beliefs. The experimentalist teacher was congruent in his experimental beliefs, and was recognized generally only by observers who shared them. In other words, "It takes one to know one."

b. The more highly an observer-judge tended to agree with experimentalism in both his educational and personal beliefs, the less experimental teaching behavior he saw. It took a teacher with very strong beliefs in experimentalism to overcome the experimentalist observer-judge's inclination to see very little experimental teaching. The converse was also found: the less an observer agreed with Dewey's philosophy, the more apt he was to "see" experimental teaching behavior wherever he looked, failing to make sharp differentiations among teachers on the TPOR.

c. Personal factors were also found to influence the behaviors seen by observers, but to lesser extent than the observers' beliefs. The older the observer the more experimental teaching behavior he saw, especially if he held conflicting educational and personal beliefs. Young female observers in high agreement with experimentalism at both levels of beliefs, were the stingiest of all in their recognition of experimental teaching behaviors.

d. The occupation of the observer seemed to influence his beliefs, or the converse, leading to the conjecture that the beliefs of an individual may well be a factor in his selection or attainment of an occupational position. Personnel of institutions of higher learning were found to hold beliefs much more in agreement with Dewey's philosophy than were public school personnel. Student teachers held beliefs somewhere between these two groups. Hence, college personnel saw much less experimental teaching behavior than did principals, supervisors, and other individuals working in the public schools.

II. Evaluations of Teacher Competence

a. The single best predictor of an evaluation score awarded to a student teacher was his observation (TPOR) score. The more his practices were seen to be in agreement with experimentalism, the higher his evaluation. However, non-experimental (low TPOR scores) did not necessarily result in low ratings.

b. The beliefs of the observer-judge also influence the ratings he gave teachers. The higher his belief scores, the lower the ratings given by him. This effect was modified if he observed a student teacher with correspondingly high scores who evidenced experimental behavior.

c. Most observer-judges tended to be moderately in agreement with experimentalism, saw more non-experimental than experimental teaching, and yet they rated student teachers, on the whole, "very good" to "excellent." Only the observer-judges who were consistently in high agreement with Dewey's philosophy seemed willing to give "poor" ratings to teachers whose behavior was predominantly non-experimental. All other observer-judges seemed willing to accept teaching behavior which was much less experimental than they said they wanted to see and pronounce it "good."

d. The beliefs of student teachers themselves had much less influence on the ratings they received than did the beliefs of the observer-judge who gave the ratings. In general, however, teachers who were observed to use teaching practices advocated by Dewey were more likely to receive high evaluative ratings than other teachers; however, teachers need not necessarily have used such practices in order to receive high teacher competence ratings.

e. Evaluations of teacher competence were also influenced by the sex of the observer-judge and the student teacher. The ratings tended to be higher if both observer-judge and student teacher were of the same sex. Also, the older the student teacher, the higher his rating.

f. There were some specific classroom practices which most observers saw as good and rewarded with high evaluations. These teacher practices were:

1. Encouraged pupil activity and participation.
2. Emphasized subject matter of challenging quality.
3. Taught in the "hypothetical mode," engaged pupils in activities calculated to generate and test hypotheses.
4. Refrained from exercising rigid and tight control of pupil activities.
5. Motivated pupils with intrinsic value of ideas or activity, rather than extrinsic rewards or threats.
6. Encouraged pupils to evaluate their own work by relative criteria, rather than to rely upon external authority and rigid standards alike for all.
7. Organized learning around a predetermined curriculum rather than the pupil's own concerns and problems (the only teaching practice on the list which is contrary or inconsistent with Dewey's philosophy).

C. Phase III constituted a follow-up study. A sample of 100 student teachers from Phase II plus 100 "experienced" teachers were observed, evaluated, and compared. In some cases observer-judges from Phase II were used, although it was necessary to add a great many new observer-judges from the public schools in the nearly two dozen states where the Phase II subjects were traced. The major findings of this phase were as follows:

I. Observations of Teacher Behavior

a. The findings of Phase II were corroborated as to the effects of a teacher's beliefs on the classroom behavior he is seen to evidence. The teacher's beliefs again were found to have a direct relationship to his observed behavior.

In addition, it was also found that the more open-minded (as measured on the D-Scale) a teacher, the more apt he was to be observed practicing experimental procedures.

b. The more open-minded the observer-judge, the less of Dewey's experimentalism he saw exhibited in the classroom.

c. Again, the occupation of the observer-judge seemed to influence his classroom observations. Public school central office personnel found the most experimental teacher behavior; professors of education, the least.

d. Secondary teachers were generally observed to use more teaching practices advocated by Dewey than elementary teachers, although female elementary teachers who were extremely high believers demonstrated the most experimental practices of all.

e. The subject matter taught also influenced the teacher's observed behavior. Reading teachers were seen as the most experimental; mathematics and foreign language teachers, the least.

II. Evaluations of Teacher Competence

a. Again, as in Phase II, the classroom behavior of the teacher (TPOR score received) was found to be the single most important factor in predicting the evaluation given him by the observer-judge. The more the teacher was seen practicing experimental behavior, the more competent he was seen. This trend was more pronounced if the teaching took place in the elementary school.

b. The beliefs of the observer-judge were again found to influence his evaluations of teachers. The more in agreement with experimentalism the judge, the lower his evaluations tended to be. The trend was even more pronounced if the judge was also open-minded. Clearly, open-minded judges who espoused experimentalism expected competent teachers to perform experimentally.

c. Experienced teachers were given higher ratings than the beginning teachers, although no significant differences were found in their observed classroom behavior. Foreign language teachers received the highest ratings, teachers of mathematics and the language arts the lowest, other subjects falling between the extremes. Contrary to the general trend of experienced teachers receiving higher evaluations than beginning teachers, new elementary teachers received significantly higher evaluations than did experienced elementary teachers.

III. Comparison of Phase II and Phase III Data

a. No significant difference was found between Phase II and Phase III ratings, indicating that evaluative ratings given student teachers tend to be predictive of ratings they receive as first-year or beginning teachers in the public schools.

b. First-year teachers were observed to use significantly more experimental teaching practices than they did as student teachers the year before. However, the observer-judges used in Phase III were significantly less experimental in their philosophic beliefs than were those used in Phase II; therefore, it could be expected that they would "see" more experimental behavior. Thus, the apparent change in teacher behavior was probably merely a reflection of the change to a new group of observer-judges with much lower expectations with respect to experimentalism.

Conclusions

While the population studied did not constitute a "scientific" or representative sample of student teachers, teacher education institutions, public schools, teachers, and supervisory personnel in America, nevertheless, the study was broadly based and included a very large number of teacher and observer-judge subjects in a wide variety of teaching situations around the country. As conclusions, we shall list those results which seem to have considerable support within the data, both in terms of consistency across the subjects studied and in terms of the consistency of relationships among the variables themselves.

1. Teachers' beliefs seem to have some influence on both the observational descriptions and the evaluative ratings of their teaching behavior--although much less clearly than do the beliefs of the observer-judges.

2. Observer-judges' beliefs appear to strongly influence both their observational descriptions and the evaluative ratings of teaching behavior.

3. Observational descriptions of teachers' classroom behavior seem to be the most powerful single predictor of ratings of teacher competence.

4. Evaluative ratings given student teachers tend to be predictive of the evaluative ratings they receive as first-year teachers in the public schools.

5. Observer-judges associated with the public schools tend to observe more experimental teaching behavior and give higher evaluative ratings than do observer-judges from the colleges and universities.

6. Teachers who use teaching practices in agreement with John Dewey's philosophy of education generally are given higher ratings than those who do not.

7. The teaching practices which observer-judges tend to see as good and reward with high ratings are:

a. Encouraged pupil activity and participation.

b. Emphasized subject matter of challenging quality.

c. Taught in the "hypothetical mode," engaged pupils in activities calculated to generate and test hypotheses.

d. Refrained from exercising rigid and tight control of pupil activities.

e. Motivated pupils with intrinsic value of ideas or activity, rather than extrinsic rewards or threats.

f. Encouraged pupils to evaluate their own work by relative criteria, rather than to rely upon external authority and rigid standards alike for all.

g. Organized learning around a predetermined curriculum rather than the pupil's own concerns and problems (the only teaching practice on the list which is contrary or inconsistent with Dewey's philosophy).

8. It seems to be less essential for high school teachers to employ experimentalist teaching practices in order to get high ratings than for elementary school teachers.

9. Evaluation of teachers in terms of global competencies seems to lack justification; instead, teacher evaluation is relative to the complex interaction of many factors, including beliefs of both the teacher and the observer-judges, observations of classroom behavior, age, sex, experience, grade level, and subject taught.

10. A serious belief gap was found between colleges of education and the public schools, with student teachers caught in the middle. This finding undoubtedly contributes to confusion in the thought and behavior of new teachers as they make the transition from teacher-training programs to public school classrooms.

11. A serious discrepancy was found within the belief systems of teachers and teacher educators. While teachers generally agreed strongly with Dewey's philosophy in their verbalizations about which practices should be employed in teaching, they were in much less agreement concerning the philosophic beliefs underlying those practices, and consequently, failed to use those practices in the classroom.

12. Teachers observed in all phases of the study tended not to use teaching practices advocated by John Dewey, i.e., they were observed

to use many more non-experimental than experimental teaching behaviors. This represents a serious theory-practice dilemma, as they failed in many cases to use the very practices which they themselves had said they should use.

13. The observer-judges also revealed a serious theory-practice discrepancy. While they tended to believe strongly that teachers should use experimental teaching practices, they saw relatively little of such practices in their observations made for this study, yet they evaluated that teaching on the whole as "good" to "excellent." While they gave their highest ratings to experimental teachers, they tended to give generally high ratings to all teachers, regardless of the practices they used.

Recommendations

Many of the problems involved in making evaluations of classroom teaching quality have been examined in this study. Procedures by which observer-judges can evaluate teaching behavior using individual criteria identifiable in terms of measured positions on relevant beliefs scales were tested, and found reasonably feasible and practical for use on a wide-scale. The findings and experience yielded in this study suggest the following recommendations:

1. Teacher evaluation programs should be based on reliable and valid descriptions of teacher behavior obtained by systematic observations. Ratings of teacher competence are meaningless unless related to such descriptive information and the value positions of observer-judges.

2. The observational instruments used in this study merit further use in gathering descriptive data for teacher evaluations, but should be supplemented by a number of additional observational systems based on different and differing educational theories. While Dewey's philosophy of experimentalism was shown to be a relevant dimension for measuring teaching behavior for evaluative purpose, it was also abundantly clear that it by no means accounts for all that can or should happen in classrooms.

3. In contrast to the procedures used in this study (in order to study sources of unreliability and invalidity of observers' use of observational tools), it is essential that future observers be carefully and thoroughly trained in the use of the observation systems in order to establish the highest possible standards of reliability and validity for the descriptive data obtained with them.

4. Complete objectivity in obtaining descriptions of classroom teaching behavior does not seem possible. Therefore, it is essential to identify the beliefs of the observers which influence their observational perceptions of behavior in order to take them into account in the interpretation of obtained scores. Whenever possible an observational instrument should have its theoretical and empirical relationships established with some valid and reliable attitude or belief scale in

order to facilitate the identification and control of observer bias.

5. Bias of the observer-judge need not be eliminated, but merely identified and taken into account. If evaluation of teaching competence is to be fair, the legitimate and legal differences of opinion or belief with respect to educational purposes or philosophy should be permitted and provided for within the evaluative process. This can be done, providing:

a. A number of observational instruments are used which represent a cross-section of differing educational philosophies.

b. A number of different observer-judges are used, each holding differing educational beliefs or philosophies.

c. Repeated observations of a teacher's behavior are made over time by a number of different observer-judges, using several different observational systems for which they have received thorough training.

6. However, the training of observer-judges should be directed primarily toward overcoming invalidity of observations which results from the strong intrusion of observer beliefs on observational scores. High reliability coefficients are fairly easy to achieve, even without training, but these can be deceiving. For example, observers who do not agree with Dewey's philosophy reliably observe many more Dewey-like teaching practices (i.e., they tend to give much higher TPOR scores) than do observers who are more knowledgeable of and sympathetic with Dewey's philosophy. From the experimentalist viewpoint, the observations of these "non-believers" are not valid. Training should attempt to reduce this source of invalidity, either by changing the beliefs of the observers or by removing observers who fail to learn to use the observational instrument from a point of view sympathetic to its theoretical framework. These "unchangeable" or "removed" observers should not be eliminated from the observer-judge team, but merely switched to another observational system--one based on a philosophy with which the observer is more sympathetic and knowledgeable. Thus, we can achieve valid use of a particular instrument, without eliminating desirable diversity in educational beliefs from the overall program of systematic observation and evaluation of teaching behavior.

7. Teacher education institutions should place greater emphasis in their programs on bridging the theory-practice gap. Teacher education programs seem to have been based on the assumption that all one has to do is to tell or show prospective teachers what to do and they will do it. Then, we wring our hands and shake our heads when they don't do it. The fallacy of this assumption is that it fails to realize that beliefs must be congruent with behavior, and that fundamental philosophic beliefs must be congruent with educational philosophy in order to establish congruency of teaching theory and practice.

Observer-judges in teacher education tend to hold underlying philosophic beliefs which run counter to the kinds of teaching behaviors they believe are good. When these two levels of belief are in conflict, behavior tends to agree with the more fundamental philosophic beliefs. This being the case, they have no alternative but to see behavior as being different than it is, their own as well as others, because they want or need to identify themselves with what they have been taught (in education courses) to believe is "good" behavior. For example, almost everybody connected with education believes in providing for individual differences among students, yet very few ever do it. Therefore, they are anxious and willing to see almost anything as providing for individual differences--thus, saving themselves and the teachers they observe and judge.

Training in and utilization of instruments for systematic observation of classroom behavior holds much promise for exposing and correcting the self-delusions which characterize the theory-practice gap in both pre-service and in-service teacher education programs.

8. State Departments of Education, in partnership with teacher education institutions and local public school systems, should establish programs of teacher evaluation to serve as a basis for both initial and continuing certification of teachers. Correction of the sources of unreliability and invalidity outlined in earlier recommendations, plus implementation of the procedures suggested in recommendation number six should serve as the basis for future teacher evaluation programs. Teacher evaluation is not an end in itself, nor is its value limited to certification purposes. The recommended procedures also can be useful in developing programs of staff, utilization and differentiation, improvement of instruction, development of curricular objectives, and the development of curriculum materials and instructional procedures which are congruent with those objectives.

APPENDIX A

TEACHER COMPETENCE RESEARCH PROJECT

UNIVERSITY OF FLORIDA
COLLEGE OF EDUCATION
GAINESVILLE, FLORIDA 32601

Your name ______________________________ Date ______________________________
(Last) (First) (Month) (Day) (Year)

Institutional affiliation ______________________________ ______________________________
(City) (State)

Position (if observer) ______________________________

Check if you are

(1) ____student teaching or interning

(2) ____observing

Year in School (if student teacher) ____________________

STUDY OF BELIEFS

This study of values contains three sections: (Part I) the Personal Beliefs Inventory involves basic philosophic questions, (Part II) the Teacher Practices Inventory involves educational questions, and (Part III) the Personal Opinion Questionnaire involves social and personal questions. THERE ARE NO "RIGHT" OR "WRONG" ANSWERS TO ANY OF THESE QUESTIONS. They are questions upon which people have legitimately different points of view. We want to know only what you yourself believe about such things.

As a matter of policy, your responses to the items in this study are confidential. TO MAINTAIN YOUR ANONYMITY THIS COVER SHEET WILL BE DISCARDED PRIOR TO PROCESSING YOUR RESPONSES. Therefore, place your name only in the space provided on this cover sheet.

Begin your answers in the following page. Answer every item for all three parts. There are no time limits. However, do not spend a lot of time puzzling over responses to items which may give you pause. Your first or immediate reaction is what we want.

Part I

PERSONAL BELIEFS INVENTORY

Form A - B

This is a study of what people believe about a number of basic philosophical questions. The best answer to each statement below is your personal belief. Many different and opposing points of view are presented here. You will find yourself believing some of the statements, not believing some, and uncertain about others. Whether you believe or do not believe any statement, you can be sure that many people feel the same as you do.

Mark each statement in the left margin by writing 1, 2, 3, or 4, 5, 6, depending on how you feel in each case.

1: I AGREE VERY MUCH
2: I AGREE ON THE WHOLE
3: I AGREE A LITTLE
4: I DISAGREE A LITTLE
5: I DISAGREE ON THE WHOLE
6: I DISAGREE VERY MUCH

_____ 1. Change is a basic characteristic of nature, and man has some measure of control over this change by using his intelligence.

_____ 2. Knowledge is truth to be accepted, held, and treasured for its own sake.

_____ 3. A statement of fact may be both true and untrue depending on the standpoints and conditions of the observations.

_____ 4. To know something is to know the inner nature of things, i.e., as they really are prior to investigation

_____ 5. Man doesn't have a "spirit" which is separable from his body and the material world.

_____ 6. Questions of value and moral judgment ought to be open to experimentation and scientific inquiry.

_____ 7. All "truths" are relative.

_____ 8. Man gains knowledge by having things impressed upon his mind.

_____ 9. Truth exists ready-made somewhere; the task of the scholar is to find it.

_____ 10. Practice is subordinate to knowledge, merely a means to it.

_____ 11. Learning is an application of mental powers to things to be known.

_____ 12. Man's destiny is in the hands of a supernatural power.

_____ 13. The mind is a group of "contents" which come from having certain material presented to it.

_____ 14. "Mind" is purely intellectual and cognitive; bodily activity is an irrelevant and intruding physical factor.

_____ 15. The ends and laws which should regulate human conduct have been determined by the superior intelligence of an ultimate Being.

_____ 16. Knowledge is the sum total of what is known, as that is handed down by books and learned men.

_____ 17. What something may be when totally independent of any observer or frame of reference is a scientifically meaningless question.

_____ 18. The mind is formed from without, as one molds and shapes a piece of clay.

_____ 19. Man's primitive impulses are neither good nor evil, but become one or the other according to the objects for which they are employed.

_____ 20. There is no spiritual realm which lies beyond man's experience in the natural world.

_____ 21. What is morally right and wrong ought to be decided on warranted evidence--the findings of empirical science.

_____ 22. Knowledge is the result of theoretical insight on the part of scholars.

_____ 23. There can be no final, absolute ends to which all men aspire.

_____ 24. The mind turns outward to truth; the emotions turn inward to considerations of personal advantage and loss.

_____ 25. The use of the scientific method can be extended to solve the problems of men in the area of values and moral judgments.

_____ 26. Man is capable of managing his own destiny in an understandable and predictable natural world.

_____ 27. The mind possesses faculties for remembering, imagining, reasoning, willing, and so forth, which are developed by exercise and discipline.

_____ 28. What is right and good at one time and place may not be right and good for all times and places.

_____ 29. You can never prove that any fact is unconditionally true.

_____ 30. The senses and muscles are merely external inlets and outlets of the mind.

_____ 31. Man's destiny is determined by circumstances of nature which are beyond his control.

_____ 32. Knowledge is artificial and ineffective in the degree in which it is merely presented as truth to be acquired and possessed for its own sake.

_____ 33. Man's choices are good only if they prove successful in helping him live with some degree of security and equilibrium in the world of nature.

_____ 34. Reaching a condition in which there were no more problems would be the ideal life.

_____ 35. In the absence of a moral code supported by absolute authority, bodily appetite and passion overpowers intelligence.

_____ 36. Questions of value and moral judgment ought to be open to experimentation.

_____ 37. Learning is the sum of impressions made on the mind as a result of presentation of the material to be known.

_____ 38. Nothing is or can be unchanging, absolutely certain.

_____ 39. The nature of a thing is determined by what it does, or can be used for; it is what it becomes with intelligent use.

_____ 40. Questions of values and morals should be taken out of their traditional supernatural setting and put in a naturalistic setting.

Part II

TEACHER PRACTICES INVENTORY

Form A - B

This is a study of what people believe is good teaching. Each statement below describes teacher practice--something a teacher might do in a classroom. Many different and opposing kinds of teacher practices are presented here. As you read these statements, you will find yourself agreeing with some, disagreeing with some, and uncertain about others. The best answer to each statement is your personal belief or opinion.

Mark each statement in the left margin by writing 1, 2, 3, or 4, 5, 6, depending on how you feel in each case.

1: I AGREE VERY MUCH	4: I DISAGREE A LITTLE
2: I AGREE ON THE WHOLE	5: I DISAGREE ON THE WHOLE
3: I AGREE A LITTLE	6: I DISAGREE VERY MUCH

_____ 1. Gives students opportunity to select facts and information which they consider appropriate to the question.

_____ 2. Usually has all students working on the same page of the same book at the same time.

_____ 3. Makes students emphatically aware that they are here to study and learn.

_____ 4. Once work has begun, insists that students remain in their places and concentrate on the task at hand.

_____ 5. Asks the kind of questions that students should be able to answer if they have studied the lesson.

_____ 6. Makes a direct presentation of the subject matter to be covered.

_____ 7. Permits students to go ahead with plans based on foresight, observation, and consideration of several alternatives--even when sure their judgment is mistaken.

_____ 8. Makes "doing something" with a thing, rather than the thing itself, the center of students' attention.

_____ 9. Focuses attention on what the students do or say, rather than on what the teacher does or says.

_____ 10. Makes the acquisition of knowledge and skills the center of students' attention and effort.

_____ 11. Has students compare the value of alternative courses of action and pass judgment on their relative desirability.

_____ 12. When one student fails to answer a question, asks another student to supply the correct answer.

_____ 13. Encourages students to suggest what might be done--to make "hypothetical leaps" into the unknown or untested.

_____ 14. Encourages students to put their suggestions to a test with such remarks as "You'll never know unless you try it."

_____ 15. Tells students where to start and what to do to accomplish the task at hand.

_____ 16. Organizes learning around questions posed by the teacher or the textbook.

_____ 17. Faithfully follows a planned schedule in order to get in the number of minutes each week allotted to each subject in the curriculum.

_____ 18. Gives students a wide choice in how they answer questions.

_____ 19. Provides a model to show students exactly what their work should be like when it is finished.

_____ 20. Gives students a free rein in devising and inventing proposals for what might be done to clear up troublesome situations.

_____ 21. Engages students in dramatizations, music, art, and other creative activities.

_____ 22. Uses a set standard to judge the work of all students in the class.

_____ 23. Insists that students face up to the realities of unpleasant predicaments and plights they get themselves into.

_____ 24. Accepts material in the approved textbook as a reliable measure for the appropriateness of information brought in by students from other sources.

_____ 25. Lets students become involved in ugly or distressing aspects of subjects.

_____ 26. Frequently asks students to choose among several alternatives.

_____ 27. Sticks to questions which can be answered by looking in the textbook or other references readily available in the school.

_____ 28. Limits physical activities to the gym or the playground.

_____ 29. Asks students to work on their own problems, rather than something made a problem only for the purpose of conveying instruction in some school subject.

_____ 30. Gives students a chance to discover by experiencing actual effects whether their choice of this rather than that idea was a judicious one.

_____ 31. Urges students to put everyday things to uses which have not occurred to others.

_____ 32. Gives students a number of starting places and a number of different ways of getting at what is to be done.

_____ 33. Provides approximately the same materials for each student in the class.

_____ 34. Shows students the most economical and efficient way to get a job done, and expects them to do it pretty much that way.

_____ 35. Allows students to move freely about the room while engaged in purposeful activity.

_____ 36. Quickly tells students whether their answers are "right" or "wrong."

_____ 37. Calls for the undivided attention of the group and scolds those who do not respond.

_____ 38. Asks the students to help decide when questions have been satisfactorily answered.

_____ 39. Encourages students to adventure into "deep water," to tackle problems that appear to be "over their heads."

_____ 40. Motivates students to greater intellectual effort by rewarding them with grades, marks, prizes, or privileges.

Part III

PERSONAL OPINION QUESTIONNAIRE

The following is a study of what the general public thinks and feels about a number of important social and personal questions. The best answer to each statement below is your personal opinion. We have tried to cover many different and opposing points of view; you may find yourself agreeing strongly with some of the statements, disagreeing just as strongly with others, and perhaps uncertain about others; whether you agree or disagree with any statement, you can be sure that many people feel the same as you do.

Mark each statement in the left margin according to how much you agree or disagree with it. Please mark every one.

Write 1, 2, 3, or 4, 5, 6, depending on how you feel in each case.

1: I AGREE VERY MUCH
2: I AGREE ON THE WHOLE
3: I AGREE A LITTLE

4: I DISAGREE A LITTLE
5: I DISAGREE ON THE WHOLE
6: I DISAGREE VERY MUCH

_____ 1. There is so much to be done and so little time to do it in.

_____ 2. Of all the different philosophies which exist in this world there is probably only one which is correct.

_____ 3. To compromise with our political opponents is dangerous because it usually leads to the betrayal of our own side.

_____ 4. Taking the Bible as a whole, one should regard it from the point of view of its beautiful mythology and literary style rather than as a spiritual revelation.

_____ 5. The present is all too often full of unhappiness. It is only the future that counts.

_____ 6. In times like these it is often necessary to be more on guard against ideas put out by people or groups in one's own camp than by those in the opposing camp.

_____ 7. Our modern industrial and scientific developments are signs of a greater degree of civilization than those attained by any previous society, the Greeks, for example.

_____ 8. A person must be pretty short-sighted if he believes that college professors should be forced to take special loyalty oaths.

_____ 9. It is only natural for a person to be rather fearful of the future.

_____ 10. Most people just don't give a "damn" for others.

_____ 11. Good government should aim chiefly at more aid for the poor, sick, and old.

_____ 12. The highest form of government is a democracy and the highest form of democracy is a government run by those who are most intelligent.

_____ 13. The United States and Russia have just about nothing in common.

_____ 14. The educational policies of the public schools should promote the study and participation in music and fine arts.

_____ 15. Modern society would benefit more from greater knowledge of the fundamental laws of human behavior than from more concern for the rights and welfare of citizens.

_____ 16. The educational policies of the public schools should undertake to increase the practical value of courses.

_____ 17. The aim of churches at the present time should be to bring out altruistic and charitable tendencies rather than to encourage spiritual worship and a sense of communion with the highest.

_____ 18. A person who thinks primarily of his own happiness is beneath contempt.

_____ 19. A group which tolerates too much difference of opinion among its own members cannot exist for long.

_____ 20. Man on his own is a helpless and miserable creature.

_____ 21. When it comes to differences of opinion in religion we must be careful not to compromise with those who believe differently from the way we do.

_____ 22. If given the chance I would do something of great benefit to the world.

_____ 23. As a branch of study theology ultimately will prove more important to mankind than mathematics.

_____ 24. It is better to be a dead hero than to be a live coward.

_____ 25. In this complicated world of ours the only way we can know what's going on is to rely on leaders or experts who can be trusted.

_____ 26. The educational policies of the public schools should stimulate the study of social problems.

_____ 27. It's the fellow travellers or Reds who keep yelling all the time about Civil Rights.

_____ 28. Abraham Lincoln should be judged as contributing more to the progress of mankind than Aristotle.

_____ 29. I'd like it if I could find someone who would tell me how to solve my personal problems.

_____ 30. While I don't like to admit this even to myself, my secret ambition is to become a great man, like Einstein or Beethoven or Shakespeare.

_____ 31. In the history of mankind there have probably been just a handful of really great thinkers.

_____ 32. The main thing in life is for a person to want to do something important.

_____ 33. Assuming that one had sufficient ability, it would be better to be a banker than a politician.

_____ 34. There are two kinds of people in this world: those who are for the truth and those who are against the truth.

_____ 35. If a man is to accomplish his mission in life it is sometimes necessary to gamble "all or nothing at all."

_____ 36. My blood boils whenever a person stubbornly refuses to admit he's wrong.

_____ 37. There are a number of people I have come to hate because of the things they stand for.

_____ 38. A man who does not believe in some great cause has not really lived.

_____ 39. It is often desirable to reserve judgment about what's going on until one has had a chance to hear the opinions of those one respects.

_____ 40. Fundamentally, the world we live in is a pretty lonesome place.

_____ 41. Even though freedom of speech for all groups is a worthwhile goal, it is unfortunately necessary to restrict the freedom of certain political groups.

_____ 42. It is more important for a child to secure training in religion than in athletics.

_____ 43. It is more important for modern leaders to bring about the accomplishment of practical goals than to encourage greater interest in the rights of others.

_____ 44. In a discussion I often find it necessary to repeat myself several times to make sure I am being understood.

_____ 45. A person who gets enthusiastic about too many causes is likely to be a pretty "wishy-washy" sort of person.

_____ 46. You just can't help but feel sorry for the person who believes that the world could exist without a Creator.

_____ 47. Good government should aim chiefly at establishing a position of prestige and respect among nations.

_____ 48. Once I get wound up in a heated discussion I just can't stop.

_____ 49. In a heated discussion I generally become so absorbed in what I am going to say that I forget to listen to what the others are saying.

_____ 50. It is justifiable for great artists, such as Beethoven, Wagner and Byron to be selfish and negligent of the feelings of others.

_____ 51. In the long run the best way to live is to pick friends and associates whose tastes and beliefs are the same as one's own.

_____ 52. One should develop one's chief loyalties toward one's occupational organization and associates.

_____ 53. A person must be pretty stupid if he still believes in differences between the races.

_____ 54. It is only natural that a person would have a much better acquaintance with ideas he believes in than with ideas he opposes.

_____ 55. Unfortunately, a good many people with whom I have discussed important social and moral problems don't really understand what's going on.

_____ 56. The worst crime a person could commit is to attack publicly the people who believe in the same thing he does.

_____ 57. Most of the ideas which get printed nowadays aren't worth the paper they are printed on.

_____ 58. Most people just don't know what's good for them.

_____ 59. The most important function of education is its preparation for practical achievement and financial reward.

_____ 60. It is only when a person devotes himself to an ideal or cause that life becomes meaningful.

_____ 61. Good government should aim chiefly at the development of manufacturing and trade.

_____ 62. The main object of scientific research should be the discovery of truth rather than its practical applications.

_____ 63. In times like these, a person must be pretty selfish if he considers primarily his own happiness.

_____ 64. One should develop one's chief loyalties toward one's religious faith.

_____ 65. High ideals and reverence are more desirable traits than unselfishness and sympathy.

_____ 66. Developing the mastery of a favorite skill is a more important use of leisure time than doing volunteer social or public service work.

APPENDIX B

TEACHER COMPETENCE RESEARCH PROJECT

UNIVERSITY OF FLORIDA
COLLEGE OF EDUCATION
GAINESVILLE, FLORIDA 32601

TEACHER'S CLASSROOM BEHAVIOR

Name of Teacher ______________________________ Date ________________________
(Month) (Day) (Year)

School ______________________________ ________________________
(City) (State)

Grade __________ Subject __________ Name of Observer-judge ______________

The teacher's classroom behavior is observed, evaluated, and recorded in this booklet, which contains:

I. TEACHER PRACTICES OBSERVATION RECORD

II. CLASSROOM BEHAVIOR RATING SCALE

III. OBSERVER-JUDGE'S COMMENTS

IV. TEACHER EVALUATION

V. RECOMMENDATION FOR CERTIFICATION

The observer-judge should complete I during his observation of the teacher in the classroom, II, III, IV, and V at the end or immediately following the observation.

I. TEACHER PRACTICES OBSERVATION RECORD

DIRECTIONS

The Teacher Practices Observation Record provides a framework for observing and recording the classroom practices of the teacher. Your role as an observer is to watch and listen for signs of the sixty-two teacher practices listed and to record whether or not they were observed, WITHOUT MAKING JUDGMENTS AS TO THE RELATIVE IMPORTANCE OR RELEVANCE OF THOSE PRACTICES.

There are three (3) separate 10-minute observation and marking periods in each 30-minute visit to the teacher's classroom. These are indicated by the column headings I, II, and III. During period I, spend the first 5 minutes observing the behavior of the teacher. In the last 5 minutes go down the list and place a check (✓) mark in Column I beside all practices you saw occur. Leave blank the space beside practices which did not occur or which did not seem to apply to this particular observation. Please consider every practice listed, mark it or leave it blank. A particular item is marked only once in a given column, no matter how many times that practice occurs within the 10-minute observation period. A practice which occurs a dozen times gets one check mark, the same as an item which occurs only once.

Repeat this process for the second 10-minute period, marking in Column II. Repeat again for the third 10-minute period, marking in Column III. Please add the total number of check marks recorded in columns I, II, and III for each teacher practice and record in the column headed TOT. There may be from 0 to 3 total check marks for each item.

TEACHER PRACTICES OBSERVATION RECORD

TOT	I	II	III	TEACHER PRACTICES
				1. T makes self center of attention.
				2. T makes p center of attention.
				3. T makes some thing itself center of p's attention.
				4. T makes doing something center of p's attention.
				5. T has p spend time waiting, watching, listening.
				6. T has p participate actively.
				7. T remains aloof or detached from p's activities.
				8. T joins or participates in p's activities.
				9. T discourages or prevents p from expressing self freely.
				10. T encourages p to express self freely.
				11. T organizes learning around Q posed by T.
				12. T organizes learning around p's own problem or Q.
				13. T prevents situation which causes p doubt or perplexity.
				14. T involves p in uncertain or incomplete situation.
				15. T steers p away from "hard" Q or problem.
				16. T leads p to Q or problem which "stumps" him.
				17. T emphasizes gentle or pretty aspects of topic.
				18. T emhpasizes distressing or ugly aspects of topic.
				19. T asks Q that p can answer only if he studied the lesson.
				20. T asks Q that is not readily answerable by study of lesson.

TOT	I	II	III	TEACHER PRACTICES
				21. T accepts only one answer as being correct.
				22. T asks p to suggest additional or alternative answers.
				23. T expects p to come up with answers T has in mind.
				24. T asks p to judge comparative value of answers or suggestions.
				25. T expects p to "know" rather than to guess answer to Q.
				26. T encourages p to guess or hypothesize about the unknown or untested.
				27. T accepts only answers or suggestions closely related to topic.
				28. T entertains even "wild" or far-fetched suggestion of p.
				29. T lets p "get by" with opinionated or stereotyped answer.
				30. T asks p to support answer or opinion with evidence.
				31. T collects and analyzes subject matter for p.
				32. T has p make his own collection and analysis of subject matter.
				33. T provides p with detailed facts and information.
				34. T has p find detailed facts and information on his own.
				35. T relies heavily on textbook as source of information.
				36. T makes a wide range of informative material available.
				37. T accepts and uses inaccurate information.
				38. T helps p discover and correct factual errors and inaccuracies.
				39. T permits formation of misconceptions and overgeneralizations.
				40. T questions misconceptions, faulty logic, unwarranted conclusions.

TOT	I	II	III	TEACHER PRACTICES
				41. T passes judgment on p's behavior or work.
				42. T withholds judgment on p's behavior or work.
				43. T stops p from going ahead with plan which T knows will fail.
				44. T encourages p to put his ideas to a test.
				45. T immediately reinforces p's answer as "right" or "wrong."
				46. T has p decide when Q has been answered satisfactorily.
				47. T asks another p to give answer if one p fails to answer quickly.
				48. T asks p to evaluate his own work.
				49. T provides answer to p who seems confused or puzzled.
				50. T gives p time to sit and think, mull things over.
				51. T has all p working at same task at same time.
				52. T has different p working at different tasks.
				53. T holds all p responsible for certain material to be learned.
				54. T has p work independently on what concerns p.
				55. T evaluates work of all p by a set standard.
				56. T evaluates work of different p by different standards.
				57. T motivates p with privileges, prizes, grades.
				58. T motivates p with intrinsic value of ideas or activity.
				59. T approaches subject matter in direct, businesslike way.
				60. T approaches subject matter in indirect, informal way.
				61. T imposes external disciplinary control on p.
				62. T encourages self-discipline on part of p.

II. CLASSROOM BEHAVIOR RATING SCALE

DIRECTIONS

At the end of the thirty-minute observation period the Classroom Behavior Rating Scale is to be used to record your overall impressions of both the teacher's behavior and the behavior of the pupils. The scale includes thirteen dimensions of teacher behavior and four dimensions of pupil behavior, which are represented by pairs of adjectives, such as ALOOF--APPROACHABLE, UNRESPONSIVE--RESPONSIVE, HARSH--KINDLY, and so forth. These behavior dimensions are described in detail in the Glossary which follows the scale.

To complete the Classroom Behavior Rating Scale you begin by studying the descriptive examples of the ALOOF--APPROACHABLE dimension of teacher behavior in the Glossary. Decide at which point on the continuum of behavior ranging from one to six you would rate the teacher. Circle the appropriate value on the rating scale. Proceed to the dimension UNRESPONSIVE--RESPONSIVE, study the glossary and rate the teacher. Continue in a similar manner until values on all dimensions of teacher behavior and pupil behavior have been assigned.

CLASSROOM BEHAVIOR RATING SCALE

TEACHER BEHAVIOR

RATING

ALOOF	←—— 6 5 4 3 2 1 ——→	APPROACHABLE
UNRESPONSIVE	←—— 6 5 4 3 2 1 ——→	RESPONSIVE
DIRECTIVE	←—— 6 5 4 3 2 1 ——→	NON-DIRECTIVE
OBSTRUCTIVE	←—— 6 5 4 3 2 1 ——→	FACILITATIVE
HARSH	←—— 6 5 4 3 2 1 ——→	KINDLY
IMMATURE	←—— 6 5 4 3 2 1 ——→	MATURE
INFLEXIBLE	←—— 6 5 4 3 2 1 ——→	ADAPTABLE
INSENSITIVE	←—— 6 5 4 3 2 1 ——→	SENSITIVE
NARROW	←—— 6 5 4 3 2 1 ——→	BROAD
PARTIAL	←—— 6 5 4 3 2 1 ——→	FAIR
PESSIMISTIC	←—— 6 5 4 3 2 1 ——→	OPTIMISTIC
OPINIONATED	←—— 6 5 4 3 2 1 ——→	OPEN-MINDED
SUSPICIOUS	←—— 6 5 4 3 2 1 ——→	TRUSTING

PUPIL BEHAVIOR

RATING

UNRESPONSIVE	←—— 6 5 4 3 2 1 ——→	RESPONSIVE
DEPENDENT	←—— 6 5 4 3 2 1 ——→	INITIATING
OBSTRUCTIVE	←—— 6 5 4 3 2 1 ——→	RESPONSIBLE
UNCERTAIN	←—— 6 5 4 3 2 1 ——→	CONFIDENT

GLOSSARY: CLASSROOM BEHAVIOR

TEACHER BEHAVIOR

ALOOF

a. Stiff and formal in relations with pupils.
b. Apart; removed from class activity
c. Condescending to a pupil.

APPROACHABLE

a. Accepting and warm in contacts with pupils.
b. Participated in class activity.
c. Regarded pupil as an equal.

UNRESPONSIVE

I. Aggressive
a. Hyper-active ignoring pupil purposes.
b. Deliberately disregarded pupil needs.
c. Was "on the prowl" looking for violators.
II. Submissive
a. Listless, lacked enthusiasm.
b. Unable to relate to pupils.
c. Bored by pupils.
d. Preoccupied, attention wandered.
e. Passive, sat in chair most of time; wasted time.

RESPONSIVE

a. Enthusiastic, interested in what was going on in class.
b. Reacted to requests and questions with sympathy.
c. Sensed needs of pupils and took steps to satisfy them.
d. Active in promoting pupil purposes.
e. Gave encouragement, commended effort.

DIRECTIVE

a. Told pupils each step to take so that future steps were uncertain.
b. Dictated particular work tasks and work companions of each pupil.
c. Intolerant of ideas or suggestions made by pupils; interrupted pupils; out-talked pupils.
d. Insisted on strict order at all times; commanded pupils; gave an order to be obeyed at once.

NON-DIRECTIVE

a. Sketched general steps to group goals; choice was allowed in alternative procedures.
b. Left divisions of tasks to the group and allowed members to work with companions of own choice.
c. Encouraged group discussion and decisions; exchanged ideas with pupils; asked opinions of pupils.
d. Guided pupils and made suggestions without being mandatory.

GLOSSARY: CLASSROOM BEHAVIOR (CONT.)

TEACHER BEHAVIOR

OBSTRUCTIVE

I. Aggressive

a. Was an inhibiting or repressive force on pupils.
b. Hovered; smothered pupils with excessive "help."
c. Made "snap" judgments.

II. Submissive

a. Was reluctant to make a decision; "passed the buck."
b. Let a difficult situation get out of hand.
c. Failed to give adequate help.

FACILITATIVE

a. Criticized in a constructive manner.
b. Made decisions as required by the situation.
c. Handled a difficult situation helpfully.
d. Suggested aids to learning.

HARSH

a. Hypercritical, fault-finding, abusive.
b. Ridiculed a pupil; depreciated a pupil's efforts.
c. Used threats; lost temper; was cross, permitted pupils to laugh at mistakes of others.
d. Was severe; grabbed, shook or otherwise "manhandled" a child.

KINDLY

a. Considerate of pupils.
b. Friendly.
c. Found good things to call attention to in a pupil.
d. Self-controlled.
e. If teacher expressed aggression, it was mild and adequate to the situation.

IMMATURE

I. Aggressive

a. Tended to condemn, reject and punish violators of conventional values.
b. "Took out" own aggressions on pupils.
c. Emotionally unstable; demanding, humorless.

II. Submissive

a. Disposed to glorify, to be subservient to and to remain uncritical of her superiors.
b. Unrealistic, naive.
c. Indicated envy; showed self-pity; was complaining.

MATURE

a. Able to criticize existing authority figures in a constructive way.
b. Realistic in approach.
c. Showed good common sense.
d. Emotionally stable; natural in manner; modest.
e. Possessed a good sense of humor.

GLOSSARY: CLASSROOM BEHAVIOR (CONT.)

TEACHER BEHAVIOR

INFLEXIBLE

a. Rigid in conforming to routine.
b. Failed to modify explanations or activities to meet new situations.
c. Resisted change from accustomed way of doing things.
d. Made no attempt to meet individual needs of pupils; demanded or compelled uniformity.
e. Impatient with interruptions or digressions from the usual.

ADAPTABLE

a. Flexible in meeting unusual situations.
b. Able to support change when new needs arose.
c. Individualized instruction
d. Was patient with interruptions and used digressions to further clarify ideas.

INSENSITIVE

a. Impatient with and opposed to the tenderminded; ignored non-volunteers.
b. Showed little concern with what pupils think and feel; recognized only academic achievement.
c. Not sympathetic with a pupil's failure at a task.
d. Looked upon pupil as if he were a physical object to be manipulated.

SENSITIVE

a. Showed awareness of pupil's emotional problems and needs.
b. Drew out non-volunteers.
c. Was alert to differences in individual thoughts and feelings.
d. Placed pupil's needs first.

NARROW

a. Showed traditional adherence to conventional values.
b. Showed little breadth of acceptance of pupil's dress, speech or manners, etc.
c. Took an uncritical approach to subject matter and problems.
d. Answers to pupil's questions were incomplete, inaccurate or unsatisfactory.

BROAD

a. Showed appreciation of other ways.
b. Evidenced a broad base of acceptance of pupil's dress, speech or manners, etc.
c. Approached subject matter of problems critically.
d. Gave complete, accurate, and satisfying answers to pupil's questions.

GLOSSARY: CLASSROOM BEHAVIOR (CONT.)

TEACHER BEHAVIOR

PARTIAL

a. Rejected or slighted a pupil.
b. Showed bias or prejudice (favorable or unfavorable) toward some social, ethnic or religious group.
c. Gave pupil special advantage; gave most attention to one or few pupils.
d. Criticized or corrected certain pupils excessively; alienated pupil by unfair criticism.

FAIR

a. Treated all pupils in an equitable and impartial manner.
b. Demonstrated freedom from prejudice toward any social, ethnic, or religious group.
c. Distributed attention among many pupils.
d. Could be critical without alienating pupil.

PESSIMISTIC

a. Appeared to see and call attention to mistakes and errors.
b. Was resigned, cynical.
c. Was depressed; seemed to be unhappy, irritable; frowned most of time.
d. Saw and called attention to the potentially bad; was skeptical.

OPTIMISTIC

a. Appeared to see and call attention to accomplishments and successes.
b. Saw and called attention to the potentially good.
c. Spoke of the future hopefully.
d. Was cheerful and good natured; joked with pupils occasionally.

OPINIONATED

a. Used facts in a falacious order or system; used facts out of context.
b. Mechanically tended to subsume things under rigid categories; evidenced hierarchical thinking.
c. Regarded each group as a homogeneous entity; no mention was made of exceptions.
d. Overgeneralized; was obtuse; lacked imagination in developing ideas.

OPEN-MINDED

a. Was scientific in using facts; approached problems with an "exploratory" attitude.
b. Showed evidence of imagination and independence of thought; showed initiative in taking new approach.
c. Regarded groups as being composed of individuals.
d. Called attention to exceptions.

GLOSSARY: CLASSROOM BEHAVIOR (CONT.)

TEACHER BEHAVIOR

SUSPICIOUS	TRUSTING
a. Refused or failed to trust a pupil's motives.	a. Showed confidence or faith in pupil's integrity.
b. Entertained doubts or showed lack of confidence in pupils.	b. Showed no doubts no reservations of a pupil's integrity.
c. Suspected pupils of hatching plots behind teacher's back.	c. Was permissive; allowed pupils freedom without showing anxiety, fear, or misgiving.
d. Intimated guilt of a pupil without proof.	d. Assumed a pupil to be innocent until proved guilty.

PUPIL BEHAVIOR

UNRESPONSIVE	RESPONSIVE
I. Aggressive	
a. Hyper-active, restless, engaged in "busy work."	a. Alert, enthusiastic, responded eagerly.
b. Over-anxious for teacher approval; ignored their peers.	b. Interested in what was going on; appeared immersed in their work.
c. Self-centered; unwilling to share.	c. Constructively busy; work concentratedly.
II Submissive	
a. Apathetic, listless, spiritless; participated half-heartedly; assumed a "don't care" attitude.	
b. Indifferent to teacher; conformed dully.	

DEPENDING	INITIATING
a. Relied on teacher for explicit directions; constantly referred to teacher.	a. Volunteered ideas, made plans and offered suggestions.
b. Reluctant to take the lead; followed in a rountinized way.	b. Took the lead willingly; went about duties purposefully.
c. Unable to carry on when teacher left the room; work deteriorated in absence of teacher.	c. Evidenced ability to work when teacher left the room; showed ability to solve their own problems.
d. Showed little ability to work things out for themselves; unable to proceed when initiative was called for.	d. Gave evidence of original thinking and creativity; showed curiosity.

3

GLOSSARY: CLASSROOM BEHAVIOR (CONT.)

PUPIL BEHAVIOR

OBSTRUCTIVE

I. Aggressive
a. Rude to teacher and to each other; deliberately noisy and disturbing.
b. Highly competitive; interrupted one another; demanded attention.
c. Aggressive, hostile, quarrelsome.
d. Intolerant; engaged in "Name calling," "scapegoating" or "tattling."

II Submissive
a. Refused to participate; obstinate.
b. Engaged in rebellious acts while casting sidelong glances at teacher.

RESPONSIBLE

a. Courteous to teacher and to each other; self-controlled and orderly without specific directions from the teacher.
b. Patient and considerate of others; waited their turn.
c. Receptive; cooperative in class activities.
d. Helpful, friendly, and good humored.
e. Showed respect and tolerance for one other; fair-minded.

UNCERTAIN

a. Submissive; deferred to judgment of the teacher.
b. Unsure of themselves; covered up by a show of "bravado" or swaggering; appeared embarrassed or self-conscious.
c. Afraid to try; timid or shy; less-able pupils were not responding.
d. Showed tenseness, nervous habits (nail biting, pencil chewing).

CONFIDENT

a. Spontaneous; entered freely into activities.
b. Undisturbed by mistakes; accepted criticism without embarrassment; natural in manner.
c. Willing to try new problems; spoke with assurance.
d. Relaxed and free from tenseness.

III. OBSERVER-JUDGE'S COMMENTS

Up to this point your observations have been restricted by the dimensions of the instruments provided for you. On this page you are encouraged to express yourself freely, in frames of reference of your own choosing, regarding the classroom performance you have just observed. The headings below are only suggestive. Feel free to ignore them or to cross them out and substitute your own.

Outstanding
STRENGTHS:

Outstanding
WEAKNESSES:

GENERAL
COMMENTS:

IV. TEACHER EVALUATION

On the basis of this observation how would you rank the ability of this teacher? Check the rank which best represents your evaluation of this teacher's competence.

	Outstanding			Very Good			Good			Fair			Poor			Incompetent		
1. Knowledge and Use of Subject Matter																		
2. Instructional Methods																		
3. Motivation of Pupil Interest																		
4. Discipline of Pupils																		
5. Personality Traits																		
6. Overall Teaching Competence																		

7. Degree of Autonomy. How free is this teacher in this classroom situation to teach according to his own beliefs, rather than according to the dictates of others. Circle your estimation of the degree of autonomy given this teacher.

complete autonomy	a great deal	a moderate amount	some	little	no autonomy

V. RECOMMENDATION FOR CERTIFICATION

1. Record the number of booklets you have completed on this person. If this is the first booklet, check "first."

(1)______first (2)______second (3)______third

(4)______fourth (5)______fifth

On the basis of this (these) observations, do you now feel prepared to make a conclusive recommendation for certification? If so, continue to the questions below. If not, stop. Observe this teacher again, until you feel prepared to do so.

2. Would you recommend that this person be certified to teach this subject at this grade level?

(1)______yes (2)______no

3. How sure are you of your recommendation concerning certification?

(1)______very sure (2)______somewhat sure (3)______unsure

4. Would you want this person to teach your child?

(1)______yes (2)______unsure (3)______no

APPENDIX C

FACTOR ANALYSIS

THEORETICAL FACTORS

Personal Beliefs Inventory

Item No.

FACTOR T-1 "Mind and Body"

5 Man doesn't have a "spirit" which is separable from his body and the material world.

13 The mind is a group of "contents" which come from having certain material presented to it.

14 "Mind" is purely intellectual and cognitive; bodily activity is an irrelevant and intruding physical factor.

18 The mind is formed from without, as one molds and shapes a piece of clay.

20 There is no spiritual realm which lies beyond man's experience in the natural world.

27 The mind possesses faculties for remembering, imagining, reasoning, willing, and so forth, which are developed by exercise and discipline.

FACTOR T-2 "Change and Certainty"

3 A statement of fact may be *both* true *and* untrue depending on the standpoints and conditions of the observations.

4 To know something is to know the inner nature of things, i.e., as they really are prior to investigation.

7 All "truths" are relative.

17 What something may be when totally independent of any observer or frame of reference is a scientifically meaningless question.

23 There can be no final, absolute ends to which all men aspire.

28 What is right and good at one time and place may not be right and good for all times and places.

29 You can never prove that any fact is unconditionally true.

33 Man's choices are good only if they prove successful in helping him live with some degree of security and equilibrium in the world of nature.

34 Reaching a condition in which there were no more problems would be the ideal life.

38 Nothing is or can be unchanging, absolutely certain.

39 The nature of a thing is determined by what it does, or can be used for; it is what it becomes with intelligent use.

FACTOR T-3 "Science and Morals"

6 Questions of value and moral judgment ought to be open to experimentation and scientific inquiry.

15 The ends and laws which should regulate human conduct have been determined by the superior intelligence of an ultimate Being.

21 What is morally right and wrong ought to be decided on warranted evidence--the findings of empirical science.

25 The use of the scientific method can be extended to solve the problems of men in the area of values and moral judgments.

36 Questions of value and moral judgment ought to be open to experimentation.

40 Questions of values and morals should be taken out of their traditional supernatural setting and put in a naturalistic setting.

FACTOR T-4 "Emotions and Intellect"

19 Man's primitive impulses are neither good nor evil, but become one or the other according to the objects for which they are employed.

24 The mind turns outward to truth; the emotions turn inward to considerations of personal advantage and loss.

30 The senses and muscles are merely external inlets and outlets of the mind.

35 In the absence of a moral code supported by absolute authority, bodily appetite and passion overpowers intelligence.

FACTOR T-5 "Freedom and Authority"

1 Change is a basic characteristic of nature, and man has some measure of control over this change by using his intelligence.

12 Man's destiny is in the hands of a supernatural power.

26 Man is capable of managing his own destiny in an understandable and predictable natural world.

31 Man's destiny is determined by circumstances of nature which are beyond his control.

FACTOR T-6 "Knowing and Doing"

2 Knowledge is truth to be accepted, held, and treasured for its own sake.

8 Man gains knowledge by having things impressed upon his mind.

9 Truth exists ready-made somewhere; the task of the scholar is to find it.

10 Practice is subordinate to knowledge, merely a means to it.

11 Learning is an application of mental powers to things to be known.

16 Knowledge is the sum total of what is known, as that is handed down by books and learned men.

22 Knowledge is the result of theoretical insight on the part of scholars.

32 Knowledge is artificial and ineffective in the degree in which it is merely presented as truth to be acquired and possessed for its own sake.

37 Learning is the sum of impressions made on the mind as a result of presentation of the material to be known.

Teacher Practices Inventory

FACTOR T-1 "Situation of Experience"

8 Makes "doing something" with a thing, rather than the thing itself, the center of students' attention.

9 Focuses attention on what the students do or say, rather than on what the teacher does or says.

21 Engages students in dramatizations, music, art, and other creative activities.

32 Gives students a number of starting places and a number of different ways of getting at what is to be done.

35 Allows students to move freely about the room while engaged in purposeful activity.

FACTOR T-2 "Development of Challenging Problem"

23 Insists that students face up to the realities of unpleasant predicaments and plights they get themselves into.

25 Lets students become involved in ugly or distressing aspects of subjects.

29 Asks students to work on their own problems, rather than something made a problem only for the purpose of conveying instruction in some school subject.

39 Encourages students to adventure into "deep water," to tackle problems that appear to be "over their heads."

FACTOR T-3 "Generation of Ideas"

13 Encourages students to suggest what might be done--to make "hypothetical leaps" into the unknown or untested.

18 Gives students a wide choice in how they answer questions.

20 Gives students a free rein in devising and inventing proposals for what might be done to clear up troublesome situations.

31 Urges students to put everyday things to uses which have not occurred to others.

FACTOR T-4 "Collection of Data"

1 Gives students opportunity to select facts and information which they consider appropriate to the question.

FACTOR T-5 "Development of Reasoned Hypothesis"

11 Has students compare the value of alternative courses of action and pass judgment on their relative desirability.

26 Frequently asks students to choose among several alternatives.

FACTOR T-6 "Application and Testing"

7 Permits students to go ahead with plans based on foresight, observation, and consideration of several alternatives—even when sure their judgment is mistaken.

14 Encourages students to put their suggestions to a test with such remarks as "You'll never know unless you try it."

30 Gives students a chance to discover by experiencing actual effects whether their choice of this rather than that idea was a judicious one.

FACTOR T-7 "Evaluation of Results"

38 Asks the students to help decide when questions have been satisfactorily answered.

FACTOR T-8 "Neglect of Direct Experiences"

4 Once work has begun, insists that students remain in their places and concentrate on the task at hand.

28 Limits physical activities to the gym or the playground.

37 Calls for the undivided attention of the group and scolds those who do not respond.

FACTOR T-9 "Reliance Upon Extrinsic Motivation"

5 Asks the kind of questions that students should be able to answer if they have studied the lesson.

16 Organizes learning around questions posed by the teacher or the textbook.

24 Accepts material in the approved textbook as a reliable measure for the appropriateness of information brought in by students from other sources.

27 Sticks to questions which can be answered by looking in the textbook or other references readily available in the school.

40 Motivates students to greater intellectual effort by rewarding them with grades, marks, prizes, or privileges.

FACTOR T-10 "Learning as an End in Itself"

3 Makes students emphatically aware that they are here to study and learn.

6 Makes a direct presentation of the subject matter to be covered.

10 Makes the acquisition of knowledge and skills the center of students' attention and effort.

12 When one student fails to answer a question, asks another student to supply the correct answer.

36 Quickly tells students whether their answers are "right" or "wrong."

FACTOR T-11 "Mechanical Following of an Established Method"

15 Tells students where to start and what to do to accomplish the task at hand.

17 Faithfully follows a planned schedule in order to get in the number of minutes each week allotted to each subject in the curriculum.

34 Shows students the most economical and efficient way to get a job done, and expects them to do it pretty much that way.

FACTOR T-12 "Imposes General Method on All Alike"

2 Usually has all students working on the same page of the same book at the same time.

19 Provides a model to show students exactly what their work should be like when it is finished.

22 Uses a set standard to judge the work of all students in the class.

33 Provides approximately the same materials for each student in the class.

Teacher Practices Observation Record

FACTOR T-1 "Nature of the Situation"

1 T makes self center of attention.

2 T makes p center of attention.

3 T makes some thing itself center of p's attention.

4 T makes doing something center of p's attention.

5 T has p spend time waiting, watching, listening.

6 T has p participate actively.

7 T remains aloof or detached from p's activities.

8 T joins or participates in p's activities.

9 T discourages or prevents p from expressing self freely.

10 T encourages p to express self freely.

FACTOR T-2 "Nature of the Problem"

11 T organizes learning around Q posed by T.

12 T organizes learning around p's own problem or Q.

13 T prevents situation which causes p doubt or perplexity.

14 T involves p in uncertain or incomplete situation.

15 T steers p away from "hard" Q or problem.

16 T leads p to Q or problem which "stumps" him.

17 T emphasizes gentle or pretty aspects of topic.

18 T emphasizes distressing or ugly aspects of topic.

19 T asks Q that p can answer only if he studied the lesson.

20 T asks Q that is not readily answerable by study of lesson.

FACTOR T-3 "Development of Ideas"

21 T accepts only one answer as being correct.

22 T asks p to suggest additional or alternative answers.

23 T expects p to come up with answer T has in mind.

24 T asks p to judge comparative value of answers or suggestions.

25 T expects p to "know" rather than to guess answer to Q.

26 T encourages p to guess or hypothesize about the unknown or untested.

27 T accepts only answers or suggestions closely related to topic.

28 T entertains even "wild" or far-fetched suggestion of p.

29 T lets p "get by" with opinionated or stereotyped answer.

30 T asks p to support answer or opinion with evidence.

FACTOR T-4 "Use of Subject Matter"

31 T collects and analyzes subject matter for p.

32 T has p make his own collection and analysis of subject matter.

33 T provides p with detailed facts and information.

34 T has p find detailed facts and information on his own.

35 T relies heavily on textbook as source of information.

36 T makes a wide range of informative material available.

37 T accepts and uses inaccurate information.

38 T helps p discover and correct factual errors and inaccuracies.

39 T permits formation of misconceptions and over-generalizations.

40 T questions misconceptions, faulty logic, unwarranted conclusions.

FACTOR T-5 "Evaluation"

41 T passes judgment on p's behavior or work.

42 T withholds judgment on p's behavior or work.

43 T stops p from going ahead with plan which T knows will fail.

44 T encourages p to put his ideas to a test.

45 T immediately reinforces p's answer as "right" or "wrong."

46 T has p decide when Q has been answered satisfactorily.

47 T asks another p to give answer if one p fails to answer quickly.

48 T asks p to evaluate his own work.

49 T provides answer to p who seems confused or puzzled.

FACTOR T-6 "Differentiation"

51 T has all p working at same task at same time.

52 T has different p working at different tasks.

53 T holds all p responsible for certain material to be learned.

54 T has p work independently on what concerns p.

55 T evaluates work of all p by a set standard.

56 T evaluates work of different p by different standards.

FACTOR T-7 "Motivation, Control"

57 T motivates p with privileges, prizes, grades.

58 T motivates p with intrinsic value of ideas or activity.

59 T approaches subject matter in direct, businesslike way.

60 T approaches subject matter in indirect, informal way.

61 T imposes external disciplinary control on p.

62 T encourages self-discipline on part of p.

EMPIRICAL FACTORS

Personal Beliefs Inventory

	FACTOR E-1 "Science and Morality"	Factor Loading
6	Questions of value and moral judgment ought to be open to experimentation and scientific inquiry.	.69
21	What is morally right and wrong ought to be decided on warranted evidence--the finding of empirical science.	.62
25	The use of the scientific method can be extended to solve the problems of men in the area of values and moral judgments.	.70

36	Questions of value and moral judgment ought to be open to experimentation.	.72
40	Questions of values and morals should be taken out of their traditional supernatural setting and put in a naturalistic setting.	.46

FACTOR E-2 "Mind vs. Body and Emotion"

14	"Mind" is purely intellectual and cognitive; bodily activity is an irrelevant and intruding physical factor.	.41
24	The mind turns outward to truth; the emotions turn inward to considerations of personal advantage and loss.	.39
34	Reaching a condition in which there were no more problems would be the ideal life.	.41

FACTOR E-3 "Knowledge for Its Own Sake"

2	Knowledge is truth to be accepted, held, and treasured for its own sake.	.53
32	Knowledge is artificial and ineffective in the degree in which it is merely presented as truth to be acquired and possessed for its own sake.	-.57

FACTOR E-4 "Relativity vs. Certainty"

3	A statement of fact may be both true and untrue depending on the standpoints and conditions of the observations.	-.49
7	All "truths" are relative.	-.60
29	You can never prove that any fact is unconditionally true.	-.50
31	Man's destiny is determined by circumstances of nature which are beyond his control.	.31
38	Nothing is or can be unchanging, absolutely certain.	-.53

FACTOR E-5 "Religion"

5	Man doesn't have a "spirit" which is separable from his body and the material world.	-.65
12	Man's destiny is in the hands of a supernatural power.	.70

15	The ends and laws which should regulate human conduct have been determined by the superior intelligence of an ultimate Being.	.74
20	There is no spiritual realm which lies beyond man's experience in the natural world.	-.65
26	Man is capable of managing his own destiny in an understandable and predictable natural world.	-.41

FACTOR E-6 "Nature of Learning"

8	Man gains knowledge by having things impressed upon his mind.	.51
11	Learning is an application of mental powers to things to be known.	.61
13	The mind is a group of "contents" which come from having certain material presented to it.	.52
18	The mind is formed from without, as one molds and shapes a piece of clay.	.46
27	The mind possesses faculties for remembering, imagining, reasoning, willing, and so forth, which are developed by exercise and discipline.	.54
37	Learning is the sum of impressions made on the mind as a result of presentation of the material to be known.	.63

Teacher Practices Inventory

FACTOR E-1 "Evils in Education"

2	Usually has all students working on the same page of the same book at the same time.	.46
5	Asks the kind of questions that students should be able to answer if they have studied the lesson.	.58
6	Makes a direct presentation of the subject matter to be covered.	.49
10	Makes the acquisition of knowledge and skills the center of students' attention and effort.	.45
12	When one student fails to answer a question, asks another student to supply the correct answer.	.44

15	Tells students where to start and what to do to accomplish the task at hand.	.58
16	Organizes learning around questions posed by the teacher or the textbook.	.64
17	Faithfully follows a planned schedule in order to get in the number of minutes each week allotted to each subject in the curriculum.	.53
19	Provides a model to show students exactly what their work should be like when it is finished.	.58
22	Uses a set standard to judge the work of all students in the class.	.54
24	Accepts material in the approved textbook as a reliable measure for the appropriateness of information brought in by students from other sources.	.40
27	Sticks to questions which can be answered by looking in the textbook or other references readily available in the school.	.43
33	Provides approximately the same materials for each student in the class.	.49
34	Shows students the most economical and efficient way to get a job done, and expects them to do it pretty much that way.	.55
36	Quickly tells students whether their answers are "right" or "wrong."	.48
40	Motivates students to greater intellectual effort by rewarding them with grades, marks, prizes, or privileges.	.48

FACTOR E-2 "Creativity"

21	Engages students in dramatizations, music, art, and other creative activities.	.50
26	Frequently asks students to choose among several alternatives.	.42
30	Gives students a chance to discover by experiencing actual effects whether their choice of this rather than that idea was a judicious one.	.46
31	Urges students to put everyday things to uses which have not occurred to others.	.60
32	Gives students a number of starting places and a number of different ways of getting at what is to be done.	.55

FACTOR E-3 "Hard-Nose Teacher"

3	Makes students emphatically aware that they are here to study and learn.	-.56
4	Once work has begun, insists that students remain in their places and concentrate on the task at hand.	-.54
37	Calls for the undivided attention of the group and scolds those who do not respond.	-.43

FACTOR E-4 "Pupil Activity"

8	Makes "doing something" with a thing, rather than the thing itself, the center of students' attention.	.59
9	Focuses attention on what the students do or say, rather than on what the teacher does or says.	.52

FACTOR E-5 "Tough Problem"

25	Lets students become involved in ugly or distressing aspects of subjects.	-.60
39	Encourages students to adventure into "deep water," to tackle problems that appear to be "over their heads."	-.51

FACTOR E-6 "Testing Hypotheses"

13	Encourages students to suggest what might be done--to make "hypothetical leaps" into the unknown or untested.	.49
14	Encourages students to put their suggestions to a test with such remarks as "You'll never know unless you try it."	.54

Teacher Practices Observation Record

FACTOR E-1 "Correct Subject Matter"

29	T lets p "get by" with opinionated or stereotyped answer.	.55
37	T accepts and uses inaccurate information.	.60
39	T permits formation of misconceptions and overgeneralizations.	.56

43	T stops p from going ahead with plan which T knows will fail.	.41

FACTOR E-2 "Rigidity of Study Topics--Teacher Control of Subject Matter"

21	T accepts only one answer as being correct.	.63
23	T expects p to come up with answer T has in mind.	.71
25	T expects p to "know" rather than to guess answer to Q.	.51
27	T accepts only answers or suggestions closely related to topic.	.55
31	T collects and analyzes subject matter for p.	.59
33	T provides p with detailed facts and information.	.57

FACTOR E-3 "Individualization of Instruction"

51	T has all p working at same task at same time.	.60
52	T has different p working at different tasks.	-.61
54	T has p work independently on what concerns p.	-.41

FACTOR E-4 "Rigid Standard for All"

47	T asks another p to give answer if one p fails to answer quickly.	.53
49	T provides answer to p who seems confused or puzzled.	.40
53	T holds all p responsible for certain material to be learned.	.48
55	T evaluates work of all p by a set standard.	.56

FACTOR E-5 "Subject Matter Quality"

13	T prevents situation which causes p doubt or perplexity.	-.46
19	T asks Q that p can answer only if he studied the lesson.	.48
20	T asks Q that is not readily answerable by study of lesson.	-.50

26	T encourages p to guess or hypothesize about the unknown or untested.	-.56
32	T has p make his own collection and analysis of subject matter.	-.46
35	T relies heavily on textbook as source of information.	.52
36	T makes a wide range of informative material available.	-.52

FACTOR E-6 "Generation and Testing of Hypotheses"

16	T leads p to Q or problem which "stumps" him.	.45
24	T asks p to judge comparative value of answers or suggestions.	.63
28	T entertains even "wild" or far-fetched suggestion of p.	.47
30	T asks p to support answer or opinion with evidence.	.68
38	T helps p discover and correct factual errors and inaccuracies.	.49
40	T questions misconceptions, faulty logic, unwarranted conclusions.	.61
44	T encourages p to put his ideas to a test.	.46
46	T has p decide when Q has been answered satisfactorily.	.58
48	T asks p to evaluate his own work.	.61
56	T evaluates work of different p by different standards.	.44

FACTOR E-7 "Pupil Activity"

2	T makes p center of attention.	-.48
4	T makes doing something center of p's attention.	-.47
6	T has p participate actively.	-.59
7	T remains aloof or detached from p's activities.	.40
8	T joins or participates in p's activities.	-.61
10	T encourages p to express self freely.	-.44
41	T passes judgment on p's behavior or work.	-.43

FACTOR E-8 "Direct-Indirect"

59	T approaches subject matter in direct, businesslike way.	-.52
60	T approaches subject matter in indirect, informal way.	.62